

Adiel Mittmann

## **ESCANSÃO AUTOMÁTICA DE VERSOS EM PORTUGUÊS**

Tese submetida ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Ciência da Computação. Orientador: Prof. Dr. rer. nat. Aldo von Wangenheim.

Florianópolis  
2016

Ficha de identificação da obra elaborada pelo autor através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Mittmann, Adiel

Escansão Automática de Versos em Português / Adiel Mittmann ; orientador, Aldo von Wangenheim - Florianópolis, SC, 2016.

300 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Ciência da Computação.

Inclui referências

1. Ciência da Computação. 2. Escansão automática. 3. Transcrição fonética. 4. Versificação. I. Wangenheim, Aldo von. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Computação. III. Título.

Adiel Mittmann

## ESCANSÃO AUTOMÁTICA DE VERSOS EM PORTUGUÊS

Esta tese foi julgada adequada para obtenção do título de doutor e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Florianópolis, 24 de outubro de 2016.

---

Prof<sup>a</sup>. Carina Friedrich Dorneles, Dr<sup>a</sup>.  
Coordenadora do Programa

### **Banca Examinadora:**

---

Prof. Aldo von Wangenheim, Dr. rer. nat.  
Universidade Federal de Santa Catarina  
Orientador

---

Prof. Thiago Alexandre Salgueiro Pardo, Dr.  
Universidade de São Paulo

---

Prof. Carlos Alberto Antunes Maciel, Dr.  
Université de Nantes

---

Prof<sup>a</sup>. Márcia Azevedo de Abreu, Dr<sup>a</sup>.  
Universidade Estadual de Campinas

---

Prof. Angelo Conrado Loula, Dr.  
Universidade Estadual de Feira de Santana

---

Prof. Mauro Roisenberg, Dr.  
Universidade Federal de Santa Catarina



## **AGRADECIMENTOS**

À Nataliya, por tanto carinho e paciência. À família, pelo apoio e confiança. Ao Alckmar, à Isabela, à Samanta e à Livia, sem cuja ajuda nada teria sido feito. Ao Maciel, ao Étienne, ao Laurent e ao Damon, pelo meu período no BCL. Ao Aldo, ao Roberto e ao PPGCC, pela oportunidade de fazer pesquisa. Ao CNPq, por ter me enviado à França. Aos membros da banca, por enriquecerem o trabalho com perguntas e sugestões.



1	2	3	4	5	6	7	8	9	10	11
A-	des-	te, ben-	de-	ca-	syl-	la-	bi,	quo-	t es-	tis
om-	ne-	s un-	di-	que,	quot-	quo-	t es-	ti-	s om-	nes.
Vin-	de, ben-	de-	cas-	sí-	la-	bos,	vós	quan-	tos	sois,
De	to-	dos	os	la-	dos	vin-	de em	pro-	fu-	são.





## RESUMO

Esta tese apresenta o Aoidos, um sistema capaz de escandir versos em português. Os versos passíveis de serem analisados pelo sistema são os métricos, aqueles para os quais o número de sílabas constitui uma característica fundamental. A partir de um poema, o sistema processa os versos palavra a palavra, encontrando sua sílaba tônica e realizando a divisão silábica. O sistema concatena então todas as palavras de um verso numa só unidade e obtém uma lista de possíveis pronúncias para o verso através da aplicação de diversas regras, realizando ou não fenômenos como sinéreses e sinalefas. As pronúncias encontradas para todos os versos de um poema são analisadas de forma conjunta para determinar a métrica do poema, ou seja, a quantidade de sílabas que cada um dos versos possui. Agrupando os versos de cada tamanho, o sistema encontra os padrões rítmicos e, dentre as pronúncias de cada verso, escolhe aquela que mais adere ao ritmo. Para validar o sistema, foram processados dez *corpora* compostos por autores diferentes e produzidos entre os séculos XVI e XX, somando mais de 100.000 versos. Através da conferência manual de 1.080 versos aleatoriamente selecionados (1% do total), averiguou-se que a taxa média de acerto do sistema foi de 99,0%, variando entre os *corpora* de 97,5 a 100,0%.

**Palavras-chave:** Escansão automática. Transcrição fonética. Versificação.



## ABSTRACT

This thesis presents a system capable of scanning verses in Portuguese. The system aims to scan metrical verses, those for which the number of syllables is a fundamental feature. Given a poem, the system first processes individual words, finding the position of their stress and dividing them into syllables. Words are then concatenated and the system finds a list of possible pronunciations for each verse, using a set of rules in order to apply phenomena such as synaeresis or synaloepha. The set of pronunciations for all verses in the poem are analyzed together in order to determine the poem's meter, that is, the number of syllables its verses should contain. By grouping together all verses of a given length, the system finds rhythmic patterns and chooses, for each verse, the pronunciation that most closely matches the pattern. In order to validate the system, it was applied to ten corpora, from different authors and whose production dates range from the 16th to the 20th century, for a total of more than 100,000 verses. By manually assessing the correctness of the scansion produced for 1.080 randomly selected verses (1% of the total), the accuracy rate was ascertained to be 99,0%, ranging from 97,5 to 100,0% among the different corpora.

**Keywords:** Automatic scansion. Phonetic transcription. Versification.



## SUMÁRIO

<b>1. Introdução</b>	<b>1</b>
1.1. Contextualização . . . . .	1
1.2. Proposta . . . . .	3
1.3. Motivação . . . . .	5
1.4. Objetivos . . . . .	6
1.5. Convenções . . . . .	6
1.6. Organização . . . . .	9
<b>2. Versificação</b>	<b>11</b>
2.1. O verso . . . . .	12
2.1.1. Metro . . . . .	12
2.1.2. Ritmo . . . . .	15
2.1.3. Tipologia . . . . .	16
2.1.4. Metaplasmos . . . . .	19
2.2. A estrofe . . . . .	24
2.3. O poema . . . . .	25
<b>3. Trabalhos relacionados</b>	<b>27</b>
3.1. Português . . . . .	28
3.1.1. LuCas e SAEP . . . . .	28
3.2. Espanhol . . . . .	31
3.2.1. Colors of Poetry . . . . .	32
3.2.2. O sistema de Gervás . . . . .	33
3.2.3. O sistema de Navarro-Colorado . . . . .	35
3.3. Italiano . . . . .	37
3.3.1. O sistema de Robey . . . . .	38
3.4. Francês . . . . .	39
3.4.1. Métromètre . . . . .	40
3.4.2. Anamètre . . . . .	43
3.5. Provençal . . . . .	44
3.5.1. O sistema de Rainsford e Scrivner . . . . .	44
3.6. Outras línguas . . . . .	45
3.6.1. Inglês . . . . .	45
3.6.2. Alemão . . . . .	49

3.6.3.	Russo . . . . .	50
3.6.4.	Tcheco . . . . .	50
3.6.5.	Latim e grego . . . . .	52
3.7.	Considerações finais . . . . .	53
<b>4.</b>	<b><i>Corpora</i></b>	<b>55</b>
4.1.	<i>Corpora</i> lexicais . . . . .	55
4.2.	<i>Corpora</i> poéticos principais . . . . .	56
4.3.	<i>Corpora</i> poéticos secundários . . . . .	75
<b>5.</b>	<b>O sistema Aoidos</b>	<b>79</b>
5.1.	Entrada . . . . .	80
5.2.	Pré-processamento . . . . .	86
5.3.	Detecção da vogal tônica . . . . .	90
5.4.	Divisão silábica . . . . .	98
5.5.	A máquina de regras . . . . .	102
5.6.	Transcrição fonética . . . . .	110
5.7.	Gerador de elocuções . . . . .	113
5.8.	Análise métrica . . . . .	121
5.9.	Análise rítmica . . . . .	125
5.10.	Detalhes técnicos . . . . .	129
5.11.	Considerações metodológicas . . . . .	131
<b>6.</b>	<b>Experimentos</b>	<b>135</b>
6.1.	Vogal tônica . . . . .	135
6.2.	Experimento principal de validação . . . . .	139
6.2.1.	Execução . . . . .	139
6.2.2.	Resultados . . . . .	142
6.2.3.	Discussão . . . . .	147
6.3.	Experimento secundário de validação . . . . .	155
6.4.	Caso de uso: os versos de Gregório de Matos . . . . .	159
6.5.	Caso de uso: os decassílabos do <i>Uraguai</i> . . . . .	169
<b>7.</b>	<b>Conclusão</b>	<b>181</b>
	<b>Bibliografia</b>	<b>189</b>
<b>A.</b>	<b>Versos inescandíveis</b>	<b>205</b>
<b>B.</b>	<b>Regras prosódicas</b>	<b>211</b>







## LISTA DE FIGURAS

5.1. Etapas do sistema e respectivos exemplos. . . . .	80
5.2. Exemplo do formato TEI . . . . .	81
5.3. Exemplo de versos com números . . . . .	82
5.4. Exemplo de verso com palavras estrangeiras . . . . .	83
5.5. Exemplo de versos emendados . . . . .	84
6.1. Gráfico dos tipos de versos na obra de Basílio da Gama . .	177
6.2. Gráfico dos tipos de versos em sete <i>corpora</i> . . . . .	180



## LISTA DE TABELAS

1.1.	Símbolos fonéticos . . . . .	7
2.1.	Classificação dos versos segundo seu número de sílabas. . .	14
2.2.	Classificação das estrofes segundo número de versos. . . .	24
3.1.	Contagem de sílabas pelos sistemas LuCas e SAEP . . . . .	29
3.2.	Erros do sistema SAEP . . . . .	30
3.3.	Sumário das ferramentas de escansão automática . . . . .	54
4.1.	<i>Corpora</i> poéticos: autores e datas . . . . .	58
4.2.	Quantidade de versos nos <i>corpora</i> . . . . .	59
4.3.	Versos, estrofes e poemas nos <i>corpora</i> . . . . .	61
4.4.	Versos repetidos, fragmentários e emendados nos <i>corpora</i> .	61
4.5.	Expansões, estrangeirismos e regularizações nos <i>corpora</i> .	62
4.6.	Versos corrigidos, faltantes, excedentes e inescandíveis . .	64
4.7.	<i>Corpora</i> poéticos secundários: autores e datas . . . . .	75
4.8.	Versos, estrofes e poemas nos <i>corpora</i> secundários . . . .	76
4.9.	Versos corrigidos e inescandíveis nos <i>corpora</i> secundários .	76
5.1.	Regras de localização da vogal tônica . . . . .	92
5.2.	Exemplo de seções para a regra C. . . . .	95
5.3.	Grupos compostos dos grafemas <i>u</i> e <i>i</i> . . . . .	101
5.4.	Exemplo de catálogo de ritmos . . . . .	130
6.1.	Acertos e erros na detecção das vogais tônicas no HOU10 .	137
6.2.	Acertos e erros na detecção da vogal tônica no <i>corpus</i> LIT30	139
6.3.	Tempo de execução do sistema ao analisar os <i>corpora</i> . . . .	140
6.4.	Resultados da verificação manual. . . . .	143
6.5.	Quantidade de regras aplicadas nos <i>corpora</i> . . . . .	148
6.6.	Quantidade de aplicações e de elocuições encontradas . . .	149
6.7.	Quantidade de <i>corpora</i> aos quais as regras se aplicaram . .	150
6.8.	Validação cruzada . . . . .	151
6.9.	Quantidade de aplicações segundo categoria das regras . .	152
6.10.	Pontuação média das aplicações segundo a categoria . . . .	153

6.11. Quantidade absoluta de versos segundo número de sílabas	154
6.12. Quantidade relativa de versos segundo número de sílabas	154
6.13. Versos de Gregório de Matos	160
6.14. Sonetos decassilábicos em espanhol de Gregório de Matos	162
6.15. Sonetos escritos por Gregório de Matos	163
6.16. Versos de Gregório de Matos: Aoidos vs. Chociay	165
6.17. Versos do <i>Uraguai</i> explicitamente classificados por Chociay	173
6.18. Versos heroicos do <i>Uraguai</i> : Aoidos vs. Chociay	175
6.19. Versos sáficos do <i>Uraguai</i> : Aoidos vs. Chociay	176
6.20. Percentuais no <i>Uraguai</i> : Aoidos vs. Chociay	177
6.21. Versos heroicos de sete <i>corpora</i> : Aoidos vs. Chociay	178
6.22. Versos sáficos de sete <i>corpora</i> : Aoidos vs. Chociay	179
6.23. Percentuais em sete <i>corpora</i> : Aoidos vs. Chociay	180

## 1. INTRODUÇÃO

1	2	3	4	5	6	7	8
Dei-	to	sí-	la-	bas	de	la-	do,
De ou-	tro	sí-	la-	ba-	s a-	rran-	co,
Tra-	ba-	lho	de-	ses-	pe-	ra-	do
E	fi-	ca o	pa-	pel	em	bran-	co.

Machado de Assis

Este capítulo apresenta o problema da escansão automática, descreve brevemente o sistema proposto e expõe os objetivos do trabalho. No fim do capítulo estão definidas algumas convenções usadas ao longo do documento.

### 1.1. Contextualização

O poeta, ao compor um poema, incorpora propositalmente nos versos certas características que podem ser percebidas quando são lidos. Duas características fundamentais ao verso em português são o número de sílabas que ele possui — seu metro — e a distribuição das sílabas tônicas em seu interior — seu ritmo. Neste contexto, escandir um verso significa dividi-lo em sílabas de forma a destacar seu metro e seu ritmo; ao verbo *escandir* corresponde o substantivo *escansão*.

Para os humanos, escandir um verso corretamente é uma tarefa natural, pelo menos para a maioria dos versos. A poesia é um caso especial da língua que falamos quotidianamente, portanto basta aplicarmos nossos conhecimentos do português e nossa cultura para capturarmos a métrica e o ritmo de um poema. Como disse Olavo Bilac, o ouvido é o melhor guia (58). Esta tese apresenta um método que emprega uma sequência de regras e análises para tentar emular a capacidade humana de escandir versos em português.

Nem todas as línguas e épocas escandem versos da mesma forma. Os gregos antigos, por exemplo, consideravam a quantidade de cada sílaba (ou seja, o tempo que se leva para pronunciá-la) como unidade básica em seus poemas, combinando sílabas longas e breves. Homero escreveu a *Ilíada* e a *Odisseia* utilizando hexâmetros iâmbicos: seis medidas de pés feitos de

uma sílaba longa e duas breves. Escandir, no caso do grego antigo, é determinar as sílabas longas e breves de um verso. O português, não distinguindo quantidades de vogais ou de sílabas, vale-se do número de sílabas que os versos contêm e da disposição das sílabas tônicas.

A tarefa de escandir versos em português, contudo, não se limita à obtenção da divisão silábica do verso, pois não existe uma única divisão; existem várias possíveis. Tome-se como exemplo o verso seguinte, dividido em sílabas de duas formas diferentes:

FIG:1.33.4.1 E tudo ajoelhado está, na noite calma:

1	2	3	4	5	6	7	8	9	10	11	12	13
E	tu-	do a-	jo-	e-	lha-	do es-	tá,	na	noi-	te	cal-	ma:
*E	tu-	do a-	joe-	lha-	do es-	tá,	na	noi-	te	cal-	ma:	

Ambas as divisões são válidas se o verso é considerado de forma isolada. Contudo, considerando-se o poema em que aparece, é possível descobrir que a forma correta só pode ser a primeira, pois todos os outros 13 versos do poema possuem a 12ª sílaba como última tônica.

Este verso exemplifica uma característica importante do português: nosso idioma muito frequentemente põe vogais em contato, tanto no interior das palavras quanto entre elas, e essas vogais podem se juntar ou se separar de várias maneiras. A letra final de uma palavra como *ajoelhado* é pronunciada de forma vocálica se considerada independentemente, porém no exemplo ela se torna semivogal quando *ajoelhado* junta-se a *está*.

Um método automático de escansão para o português, portanto, não pode encontrar uma versão final para a separação silábica de um verso considerando-o em isolamento. Se este fosse o caso, a junção das divisões silábicas de cada palavra, seguida da aplicação de algumas regras fixas, produziria a escansão correta. Como a revisão dos trabalhos relacionados mostrará, sistemas que se comportam desta maneira ao escandir versos em português e línguas próximas não atingem boas taxas de acerto.

O português, por outro lado, apresenta algumas características que facilitam a escansão. Após as reformas ortográficas da década de 40, encontrar a vogal tônica de uma palavra é um processo determinístico: dada uma palavra escrita, o leitor sabe, para a quase totalidade das palavras, onde está a vogal tônica. As regras que o leitor deve utilizar para descobrir a vogal tônica podem não ser nem muito simples e nem tão poucas, mas elas existem. Em outros idiomas, como inglês ou italiano, determinar a sílaba tônica implica consultar um dicionário em que ela esteja marcada.

Há versos que não são métricos: os versos livres não se preocupam com o número de sílabas ou com uma noção muito rígida de ritmo. Es-

ses versos estão fora do escopo desta tese. A fronteira entre versos livres e métricos, contudo, não é tão bem definida quanto se possa crer. Há poemas que, observados a uma certa distância, parecem conter versos livres: a largura visual dos versos varia muito, tanto entre estrofes quanto dentro de uma única. Um exame mais cuidadoso pode revelar, todavia, que os versos individualmente possuem um número de sílabas e um padrão rítmico compatível com poemas métricos. Há um contraste muito grande entre estes poemas com vários metros, chamados polimétricos, e aqueles em que todos os versos seguem o mesmo padrão, chamados isométricos. Nesta tese, poemas polimétricos são considerados objetos de estudo, mas aqueles que contenham versos livres não.

## 1.2. Proposta

Esta tese apresenta um sistema, chamado Aoidos, cujo objetivo é escanear automaticamente poemas métricos escritos em português de acordo com a ortografia contemporânea conforme praticada no Brasil. O nome do sistema é uma transliteração descomprometida da palavra grega antiga *ᾠιδός*, que significa aedo, cantor, poeta. O nome pode ser pronunciado seguindo o acento em grego, fazendo uma oxítone, ou seguindo a escrita em português, fazendo uma paroxítone.

O Aoidos parte de um poema escrito e, através da sucessiva aplicação de regras e algoritmos, produz uma escansão para cada verso do poema. O sistema trabalha, de forma geral, em quatro etapas: a transcrição fonética, a geração de alternativas de pronúncia para os versos, a determinação da métrica do poema e a seleção da pronúncia com a melhor adequação rítmica.

A transcrição fonética trabalha individualmente com cada palavra. Inicialmente a posição da vogal tônica (ou a ausência dela) é determinada. Em seguida, cada grafema é classificado em vogal, consoante ou semivogal, e a divisão silábica da palavra é estabelecida. Cada grafema é então convertido em um símbolo do Alfabeto Fonético Internacional através de um mapeamento que leva em consideração o contexto de cada grafema: assim, um *s* em início de palavra recebe um tratamento distinto de quando ocorre entre vogais em seu interior.

As transcrições fonéticas das palavras de um verso são concatenadas para gerar uma pronúncia para o verso como um todo. A partir dessa pronúncia inicial, regras prosódicas são aplicadas para produzir pronúncias alternativas. Tais regras correspondem a fenômenos conhecidos na

poesia, como é o caso da sinérese e da sinalefa. O número de variantes produzidas para cada verso depende, em geral, do número de vogais postas em contato e do número de maneiras que o sistema encontra para resolver cada encontro; versos simples podem possuir não mais do que uma variante, enquanto versos longos e com vários pontos de contato podem gerar centenas de alternativas.

De posse do conjunto de pronúncias alternativas para cada verso de um poema, o sistema busca determinar quantas sílabas cada verso possui. O sistema analisa algumas hipóteses, como a de que todos os versos possuem uma mesma quantidade de sílabas, ou então a de que o poema ou cada estrofe sua contenha versos de dois tamanhos. O resultado desta etapa é a atribuição a cada verso de um número de sílabas que o sistema acredita que ele tenha.

Na etapa final, os versos de mesmo tamanho são considerados em grupos, independentemente de sua localização no poema. Para cada grupo o sistema realiza uma análise rítmica para encontrar as sílabas comumente acentuadas. Entre as alternativas produzidas anteriormente para cada verso, o sistema escolhe como final aquela que possuir o número correto de sílabas e tiver a melhor correspondência com o padrão rítmico do grupo. Assim como os humanos, que, ao lerem um poema, já possuem certas expectativas sobre o ritmo dos versos, também o sistema pode ser parametrizado com um catálogo de ritmos que o auxilia na decisão de qual alternativa é a mais adequada.

Nesta proposta, o canal de comunicação entre o sistema e o poeta é estabelecido por intermédio da ortografia. É com base nela que o sistema faz a transcrição fonética, ponto de partida para as demais etapas. A transcrição, entretanto, não tem como objetivo reproduzir a pronúncia exata do poeta; ela é apenas uma representação da língua falada mais regular e mais precisa do que aquela da ortografia. Deste modo, a transcrição fonética adotada pelo sistema é bastante larga e posiciona-se próxima à ortografia; com efeito, fosse a ortografia atual mais fonética e capaz de representar sutilezas de pronúncia — como especificar que o vocábulo *cruel* deve ser pronunciado em uma ou duas sílabas? —, talvez o sistema nem mesmo precisasse de uma transcrição. Mesmo tomando as palavras do poeta sob a lente da ortografia contemporânea, o sistema mostra-se capaz de escanear versos de vários períodos e locais.

O sistema aqui apresentado é o primeiro de seu tipo para a língua portuguesa. Outras propostas para a escansão automática de versos em português existem, mas são poucas e com âmbito bastante limitado; elas



são discutidas no capítulo de revisão da literatura. Mais numerosas e ambiciosas são propostas para outros idiomas, incluindo o espanhol e o francês, que enfrentam desafios similares aos da presente tese.

### 1.3. Motivação

Têm recebido o nome de *humanidades digitais* os esforços da ciência da computação aplicados aos problemas particulares das humanidades. É desde a década de 40 que os computadores auxiliam nos afazeres das humanidades, mas foi somente no fim dos anos 90 que o termo surgiu, substituindo outros previamente utilizados (106). Inicialmente a computação era apenas uma ferramenta que agilizava etapas específicas do trabalho das humanidades, mas atualmente as tecnologias computacionais fazem parte integral do trabalho dos pesquisadores da área; a mudança do papel que a computação tem exercido é justamente o que motivou a criação de um novo termo (56). É no contexto das humanidades digitais que as ferramentas de escansão automática se encontram: a ciência da computação busca reproduzir uma atividade até então feita de forma manual pelos estilólogos, os estudiosos do verso.

Pela impossibilidade de analisar manualmente todas as obras literárias escritas num determinado período, a crítica literária costuma reduzir a discussão a um *cânone* de obras consideradas representativas ou influentes. Através de ferramentas automáticas, contudo, surge a possibilidade de extrair e processar informações com base num número de obras muito superior àquele que um humano pode ler ao longo de sua vida inteira: trata-se da chamada *leitura distante* (134). Diante de milhões de versos escritos em português, a única via factível de extrair informações estilísticas é a automática, por intermédio do computador. Ao compor um *cânone* de poemas ou poetas, o que se faz é efetivamente decidir o que será completamente ignorado, o que, dado o tamanho reduzido dos cânones, significa ignorar quase tudo (170).

Esta tese apresenta um sistema de escansão automática para a língua portuguesa, suficientemente veloz e preciso de modo a permitir que características estilísticas sejam obtidas a partir de grandes quantidades de versos. Tais características podem ser extraídas de vários estágios do sistema, revelando desde a proporção de vocábulos proparoxítonos empregados pelo poeta até os padrões rítmicos preferidos por ele. Esta automação permitirá o estudo e a comparação de poemas, poetas e períodos em grande escala, o que auxiliará na leitura distante de *corpora* poéticos.

## 1.4. Objetivos

A *pergunta de pesquisa* que esta tese busca responder é a seguinte: é possível construir um sistema capaz de produzir, para versos escritos em português, escansões compatíveis com aquelas feitas por humanos? O *objetivo geral* desta tese, portanto, é propor e validar um método que realiza a escansão de versos em português com um desempenho aceitável. Deste objetivo geral desprendem-se os *objetivos específicos*:

- Estabelecer e disponibilizar um conjunto de *corpora* poéticos em formato XML que sirva tanto para a realização dos experimentos desta tese quanto para pesquisas futuras, inclusive de terceiros;
- Propor e avaliar um algoritmo para a localização da vogal tônica de vocábulos escritos de acordo com a ortografia do português praticada no Brasil contemporaneamente, tanto para aplicações poéticas quanto para outras;
- Propor um algoritmo para a divisão silábica das palavras do português, novamente de forma não específica à poesia;
- Propor um conjunto de regras prosódicas para a escansão de versos em português que sejam compatíveis com a tradição versificatória;
- Propor um algoritmo para a detecção da métrica de poemas em português, ou seja, que determine o número de sílabas poéticas que os versos de um poema contêm;
- Propor um algoritmo para a detecção do ritmo de versos em português, isto é, que estabeleça um padrão rítmico aceitável para os versos de um poema;
- Avaliar o desempenho do sistema como um todo através da verificação manual de seus resultados, comprovando que os algoritmos propostos trabalham de forma harmônica para a obtenção de uma escansão aceitável dos versos de um poema.

## 1.5. Convenções

Nesta tese, a palavra *acento* refere-se única e exclusivamente ao fenômeno fonético; as marcações gráficas dispostas sobre as letras para alterar-lhes

Tabela 1.1.: Símbolos do Alfabeto Fonético Internacional utilizados nesta tese para a transcrição fonética do português. Os sons à esquerda são consoantes, os à direita vogais e semivogais.

Fonema	Exemplos	Fonema	Exemplos
/p/	<b>p</b> ato, <b>ap</b> ós	/a/	passo, <b>a</b> rtesanato, mesa
/b/	<b>b</b> arco, <b>trab</b> alho, <b>sob</b>	/e/	pres <b>e</b> , <b>e</b> scola
/m/	<b>m</b> arço, <b>am</b> or, <b>l</b> ama	/ɛ/	catéter, <b>maz</b> ela
/t/	<b>t</b> atu, <b>abac</b> ate	/ɪ/	<b>a</b> ve, <b>parede</b>
/d/	<b>d</b> ado, <b>abade</b>	/i/	<b>i</b> dade, <b>mín</b> imo
/n/	<b>n</b> avio, <b>panetone</b> , <b>cano</b>	/o/	<b>calor</b> , <b>polí</b> cial
/k/	<b>c</b> arne, <b>paca</b> , <b>leque</b>	/ɔ/	<b>fós</b> foro, <b>moda</b>
/g/	<b>g</b> alo, <b>prego</b> , <b>jegue</b>	/ʊ/	<b>tronco</b> , <b>labirinto</b>
/f/	<b>f</b> aca, <b>mofo</b>	/u/	<b>lua</b> , <b>mús</b> ica
/v/	<b>v</b> aca, <b>palav</b> ra	/ã/	<b>ânc</b> ora, <b>maçã</b> , <b>cão</b>
/s/	<b>s</b> ala, <b>páss</b> aro, <b>exceç</b> ão	/ẽ/	<b>ên</b> fase, <b>vagem</b> , <b>empe</b> cilho
/z/	<b>z</b> agueiro, <b>lousa</b> , <b>execuç</b> ão	/ĩ/	<b>ímp</b> ar, <b>ingl</b> ês
/ʃ/	<b>ch</b> ácara, <b>xí</b> cara, <b>paix</b> ão	/õ/	<b>ond</b> a, <b>comp</b> asso, <b>leões</b>
/ʒ/	<b>j</b> anela, <b>abajur</b> , <b>viagem</b>	/ũ/	<b>mund</b> o, <b>fund</b> ação
/r/	<b>p</b> arada, <b>cantor</b> , <b>arte</b>	/w/	<b>cacau</b> , <b>canh</b> ão, <b>pinguim</b>
/ʁ/	<b>r</b> apaz, <b>barro</b>	/j/	<b>paí</b> , <b>mãe</b>
/l/	<b>l</b> ava, <b>vela</b> , <b>alto</b> , <b>sol</b>		
/ʎ/	<b>lh</b> ama, <b>folh</b> a		
/ɲ/	<b>nh</b> oque, <b>fronh</b> a		

certas características (inclusive o acento) são chamadas de *diacríticos*. Assim, a palavra *embarcação* possui o acento e um diacrítico na última sílaba; a palavra *tabu* possui o acento na última sílaba e nenhum diacrítico; a palavra *mãozinha* possui o acento na penúltima sílaba e um diacrítico na primeira.

A transcrição fonética a que este trabalho se refere é do tipo larga, ou seja, não tem pretensões de codificar em detalhes todas as propriedades que a pronúncia de um verso deveria possuir. A transcrição fonética é dada entre barras e utiliza os símbolos do Alfabeto Fonético Internacional: a palavra *batalha* tem, nesta tese, a transcrição /ba'ta.ʎa/. Quando o contexto permitir, as barras não são usadas. Quando necessário, os pontos marcam a separação silábica e o símbolo parecido com o apóstrofo indica que a sílaba seguinte recebe o acento primário da palavra; acentos secundários não são utilizados ou marcados. O conjunto de símbolos fonéticos adotados nesta tese estão exibidos e exemplificados na Tabela 1.1. Adicionalmente, semivogais derivadas de vogais plenas são indicadas

através de um símbolo específico, como na primeira sílaba do seguinte verso, em que a vogal /a/ se transforma numa semivogal:

1    2    3    4    5    6    7    8    9    10    11  
 ANJ.233.2.1 É a|flo-|ra-|ris-|to-|crá-|ti-|ca-|da-|fes-|ta.  
 'ε̃a|flo|ra|ris|to|'kra|ti|ka|da|'fes|ta

Os versos utilizados ao longo da tese vêm todos dos *corpora* estabelecidos para este trabalho, à exceção daqueles exemplificados na revisão da literatura. As escansões apresentadas correspondem àquelas produzidas pelo sistema automaticamente; quando não o são, estão marcadas com asterisco. As referências a versos são feitas através de um código que os identifica dentro dos *corpora*, como nestes versos:

1    2    3    4    5    6    7    8    9    10    11  
 MAT:29.41.4.1 De |que|ser-|ve u-|ma|vi-|da a-|bo-|rre-|ci-|da?  
 CAM.1.2.8 Se a|tan-|to |me a-|ju-|da-|r o en-|ge-|nho|e ar-|te.

As três primeiras letras do código identificam o *corpus*, neste caso a obra completa de Gregório de Matos e *Os Lusíadas* de Camões. Alguns *corpora* encontram-se divididos em mais de um arquivo físico; neste caso, as três letras são seguidas de dois pontos e do número da parte, como no caso do primeiro verso. Todos os arquivos estão disponíveis *online* no seguinte endereço:

<https://github.com/adiel-mittmann/poemas>

Os nomes dos arquivos são compostos das três referidas letras mais o número da parte, sempre com dois dígitos. Os *corpora* que não contêm subdivisões recebem 1 como o número da parte. Os dois versos mencionados, portanto, encontram-se nos arquivos *mat29.xml* e *cam01.xml*. Os três números que aparecem depois dos dois pontos (ou depois do número da parte) correspondem, respectivamente, ao número do poema dentro do arquivo, ao número da estrofe dentro do poema, e ao número do verso dentro da estrofe. Nos arquivos XML os números estão associados ao atributo *n* dos elementos *div*, *lg* e *l*. Na versão em PDF desta tese, o código é atalho para a linha do arquivo no GitHub associada a ele.

Nas tabelas ao longo do documento, quando se considerou útil, o maior valor de uma coluna foi marcado com negrito e o menor com sublinhado. Também especialmente nas tabelas, mas ocasionalmente no texto, o símbolo ‰ foi utilizado para significar *a cada mil*. Quando se diz, por exemplo, que o número de erros relativo ao número de versos é 3‰, isto significa que há 3 erros a cada mil versos.

A tese não utiliza mais do que dois níveis de aninhamento de seções. Assim, o leitor encontrará no máximo uma subseção de seção de um

capítulo. Quando se julgou útil, um pequeno título em negrito foi posto no início do parágrafo que inicia uma discussão.

## 1.6. Organização

O restante do trabalho está organizado do seguinte modo. O Capítulo 2 busca familiarizar o leitor com algumas noções de versificação. A nomenclatura utilizada posteriormente fica, desta forma, definida neste capítulo. Leitores habituados com a terminologia e os conceitos de versificação podem tranquilamente pular o referido capítulo.

O Capítulo 3 discute os trabalhos relacionados à tese. Há de fato poucos trabalhos sobre a análise automática de versos em português, de forma que este capítulo leva em conta trabalhos em outros idiomas, especialmente aqueles próximos ao português.

O Capítulo 4 descreve em boa quantidade de detalhes todos os *corpora* utilizados nos experimentos e também no próprio desenvolvimento do sistema. Estes *corpora* não são apenas poéticos, pois as etapas iniciais do sistema (localização do acento e silabação) têm aspirações que vão além da poesia.

O Capítulo 5 apresenta o sistema de escansão automática de versos em Português, o Aoidos. Este é o capítulo que contém as principais contribuições do trabalho. Todas as etapas do sistema são descritas em detalhe.

O Capítulo 6 descreve três experimentos e dois casos de uso. O experimento principal de validação, que avalia a aplicação do Aoidos a mais de 100 mil versos, está nele descrito. O objetivo deste capítulo é validar a proposta apresentada no capítulo anterior.

O Capítulo 7, por fim, discorre acerca das conclusões, limitações e trabalhos futuros.

Existem ainda três apêndices: o Apêndice A contém os versos considerados inescandíveis durante as revisões dos *corpora*; o Apêndice B detalha todas as 159 regras prosódicas definidas para a escansão dos versos dos *corpora*; finalmente, o Apêndice C mostra todos os 1.080 versos manualmente verificados no experimento principal de validação.



## 2. VERSIFICAÇÃO

1	2	3	4	5	6	7	8
Prin-	cí-	ptos	co-	me a a-	fê-	re-	se,
A	pró-	te-	se o-	s in-	ven-	ta;	
No	mei-	o	ti-	ra a	sín-	co-	pe;
A e-	pên-	te-	se a-	cre-	scen-	ta;	
Cor-	ta	nos	fin-	s a a-	pó-	co-	pe;
Pa-	ra-	go-	ge o-	s au-	men-	ta.	

Visconde de Castilho

O objetivo deste capítulo é familiarizar o leitor com conceitos, termos e convenções necessários ao entendimento dos capítulos seguintes. As definições contidas no capítulo tomam como fundamento o que é normalmente aceito nos tratados de versificação, mas não há aqui nenhuma pretensão de fazer uma análise comparativa ou histórica; o que se quer é apresentar os conceitos de forma geral e sempre guiando-se pela tarefa à mão: a escansão. Citações são feitas apenas quando afirmações específicas estão sendo discutidas. Os versos de exemplo foram encontrados pelo próprio autor, exceto quando indicado. As obras consultadas para estabelecer as definições foram as seguintes, em ordem cronológica de publicação:

- *A Arte de Fazer Versos*, um manual bastante prescritivo escrito por Duque-Estrada (87), autor do nosso hino nacional;
- *Ritmo e Poesia*, um pequeno livro que influenciou gerações de futuros esticólogos, escrito por Proença (154);
- *Teoria Literária*, um livro bastante completo e repleto de exemplos, escrito por Tavares (166);
- *Teoria do Verso*, obra muito bem organizada e com um foco eminentemente prático, de Chociay (75);
- *Versos, Sons, Ritmos*, uma pequena obra didática, escrita por Goldstein (99);

- *Tratado de Versificação*, que é uma espécie de expansão feita sobre a obra de Proença, lançada por Mattoso (126).

O sistema de versificação de uma tradição poética firma-se nos elementos fonológicos do idioma (96). Em português, as características presentes no idioma exploradas pela poesia são várias, mas de particular importância é o acento; é com base nele que são contadas as sílabas dos versos e é a distribuição dele no interior do verso que define o ritmo. Estas características são bastante naturais ao nosso ouvido, mas não são universais. Na poesia grega antiga, nem o número de sílabas em cada verso nem a posição dos acentos — que, de qualquer modo, eram musicais e não dinâmicos — tinham valor poético; no inglês antigo, somente as sílabas tônicas eram contadas, as átonas sendo ignoradas na contagem (96). Este capítulo trata somente do material relevante à poesia em português.

## 2.1. O verso

Um *verso* é uma linha de poesia. Esta tese trabalha apenas com *versos métricos*, isto é, cujo número de sílabas tem importância fundamental no poema. Em português, este tipo de verso predominou desde o século XIII, com as cantigas d'amigo (77), até o fim do século XIX, quando perdeu o papel de protagonista para o *verso livre* (132), que não segue medidas silábicas. Este último tipo não é aqui objeto de pesquisa e portanto não é descrito. Nas subseções seguintes são apresentados conceitos relacionados ao metro, o tamanho silábico dos versos; ao ritmo, a distribuição dos acentos dentre dele; aos tipos comuns de versos; e aos fenômenos que fazem com que as sílabas e os acentos das palavras se ajustem aos requerimentos da métrica e do ritmo.

### 2.1.1. Metro

Quando um verso é pronunciado, de particular importância é a sua divisão em sílabas. Essa divisão não segue necessariamente a separação silábica dicionarizada das palavras que compõem o verso. Tanto o número de sílabas de cada palavra pode ser, e com frequência o é, aumentado ou diminuído quanto os encontros vocálicos nas fronteiras entre as palavras podem ser resolvidos de formas a acrescentar ou remover sílabas à contagem total do verso. O texto de um verso, portanto, pode muitas vezes ser dividido em sílabas de mais de uma maneira.



O número exato de sílabas de um verso métrico é normalmente determinado pelo poema em que se encontram, mas existem poemas com versos de variados tamanhos e neste caso cada verso precisa ser individualmente examinado pelo leitor que quiser escandir-lo. Quando o tamanho de um verso é conhecido de antemão — seja porque o leitor conhece previamente a métrica do poema ou porque o verso em questão não é o primeiro do poema e o leitor já tem suas expectativas —, as sílabas das palavras e os encontros de vogais são examinados pelo leitor, segundo certas regras, para que o número de sílabas esperado seja obtido. O leitor pode, se desejar, fazer a leitura de um poema sem se preocupar com considerações métricas; nesta tese, contudo, entende-se que o leitor está fazendo a escansão dos versos e portanto as questões métricas são de vital importância para ele.

As sílabas de um verso são contadas até a última tônica; aquelas localizadas além devem ser pronunciadas normalmente, mas não são contadas. Todos os versos seguintes, portanto, têm 10 sílabas:

	1	2	3	4	5	6	7	8	9	10	11	12
MAT.29.9.1.11	Mas	po-	rém	ad-	ver-	tin-	do,	que	há	de	ser	
MAG.22.9.1	E	qual	o	fim	se-	rá	da	Hu-	ma-	ni-	da-	de?
ANJ.12.81.1	O	ins-	tin-	to	de	pro-	cria-	r,	a	ân-	sia	le-
											gí-	ti-
											ma	

Esta contagem que despreza as sílabas átonas finais é chamada de *aguda* ou *francesa*. Até pelo menos fins do século XVIII (75), a versificação portuguesa contava sempre uma sílaba a mais além da tônica, baseando-se no princípio de que a maioria das palavras em português são paroxítonas e que, portanto, a maioria dos versos cuja última sílaba tônica é a 10ª contará com 11 sílabas totais. Esta última contagem chama-se *grave* ou *espanhola*. Esta diferença nos padrões de contagem significa que Camões contaria 11 sílabas nos versos de sua obra prima *Os Lusíadas*, ao passo que os teóricos hoje contam 10. Nos dois padrões a âncora para a contagem é, de qualquer modo, a última sílaba tônica, de forma que converter uma contagem de um padrão para outro é uma simples questão de adicionar ou subtrair uma unidade. Esta tese utiliza exclusivamente o padrão agudo de contagem, mesmo quando se estiver discutindo exemplos antigos.

Os versos que não possuem nenhuma sílaba além da tônica são chamados de *agudos*; os que possuem uma átona final são denominados *graves*; e àqueles que contam duas átonas após a sílaba acentuada final dá-se o nome de *esdrúxulos*. Apesar de certas formas verbais acompanhadas de pronomes formarem uma unidade acentual cuja sílaba tônica recua para além da antepenúltima, como em *perdoaram-no-lo* (126), o autor da pre-

Tabela 2.1.: Classificação dos versos segundo seu número de sílabas.

Número	Nome
1	Monossílabo
2	Dissílabo
3	Trissílabo
4	Tetrassílabo
5	Pentassílabo
6	Hexassílabo
7	Heptassílabo
8	Octossílabo
9	Eneassílabo
10	Decassílabo
11	Hendecassílabo
12	Dodecassílabo

sente não encontrou nomenclatura ou exemplos do fenômeno em fim de verso.

Os versos recebem nomes diferentes conforme seu tamanho, como mostra a Tabela 2.1. Os versos com mais de 12 sílabas são agrupados sob a denominação de *bárbaros*. Quanto maiores os versos de um poema, mais difícil é para o leitor encontrar a regularidade pretendida pelo poeta. Os maiores versos não bárbaros, por exemplo, possuem variações rítmicas específicas que ajudam o ouvido do leitor. É por este motivo que são raros versos métricos com mais de 12 sílabas: seria exigir demais do leitor ouvir padrões poéticos em versos tão longos. Em francês, por exemplo, Delente e Renault (81) afirmam que a percepção de igualdade de número de sílabas entre versos diferentes está limitada a 8 sílabas ou menos.

Os versos discutidos até aqui são chamados *simples* por serem constituídos de apenas uma unidade, mas há também os *compostos*, que são efetivamente dois ou mais versos simples justapostos. A intenção do poeta em tais casos seria unir semanticamente e temporalmente dois ou mais versos, de forma que uma unidade semântica maior esteja presente numa única linha e sem a pausa que está associada ao fim do verso. O seguinte verso, por exemplo, pode ser analisado como composto de dois pentassílabos:

1    2    3    4    5    6    7    8    9    10  
MAT:39.9.25.2 \*O|san-|to Hi-|me-|neu|o|pi-|ca-|ro|vil,

Ele a princípio pode parecer um decassílabo comum, mas existem outros versos no mesmo poema que contam 11 sílabas, como o seguinte:

1 2 3 4 5 6 7 8 9 10 11  
MAT:39.9.25.4 Co-men-do, e be-ben-do co-mo um Mu-cha-chim.

O que sucede é que cada um deles é composto de duas metades:

1 2 3 4 5 6  
\*O san-to Hi-me-neu  
\*o pí-ca-ro vil,  
\*Co-men-do, e be-ben-do  
\*co-mo um Mu-cha-chim.

As metades de um verso composto de dois menores são comumente chamadas de *hemistíquios* e a pequena pausa entre elas de *cesura*.

### 2.1.2. Ritmo

A distribuição das sílabas tônicas dentro de um verso é relevante na versificação. Além da última sílaba tônica, que dá ao verso a sua contagem de sílabas, as demais exercem um papel importante ao encaixar o verso em padrões estabelecidos na tradição poética. As posições silábicas que recebem um acento no verso são chamadas de *ictos*. Há uma distinção a ser feita entre os ictos dos versos e as sílabas tônicas das palavras que o compõem, pois nem todas as sílabas tônicas das palavras são ictos e nem todos os ictos dependem de sílabas tônicas lexicais. No seguinte verso, por exemplo, ocorre o choque entre duas vogais tônicas na 5ª e na 6ª posição, mas somente uma delas pode corresponder ao icto:

1 2 3 4 5 6 7 8 9 10 11  
CAM.1.1.1.2 Que-da o-ci-den-tal prai-a Lu-si-ta-na,

Já no seguinte exemplo o pronome relativo *que*, que normalmente é átono, torna-se tônico e um icto do verso por vontade do poeta:

1 2 3 4 5 6 7  
MAT:1.15.1.9 Que-ro-fi-nal-men-te, que

Por ser o conceito de escansão central à esta tese, mais atenção será dada à sua definição. Dentre os tratados de versificação consultados, Duque-Estrada (87), Proença (154) e Chociay (75) não definem o termo, apesar de que este último utiliza tanto *escansão* quanto *escandir*. Tavares (166) diz que escansão “é a contagem dos sons dos versos”, alertando imediatamente que “as sílabas métricas diferem das sílabas gramaticais em alguns aspectos”, o que esclarece que os sons com os quais o autor se preocupa são as sílabas. Escandir é, portanto, contar as sílabas métricas, que não coincidem necessariamente com as gramaticais. Goldstein (99) afirma que escandir “significa dividir o verso em sílabas poéticas”, igualmente chamando a atenção para o fato de que “nem sempre as sílabas poéticas corres-

pondem às sílabas gramaticais”. Para contar as sílabas, portanto, é preciso dividir os versos nas sílabas que o compõem. Mattoso (126) inclui em seu tratado um glossário do qual consta a entrada *escansão*: “Procedimento estilológico para verificar a correção do verso sob o ponto de vista métrico. Escandir um verso é dividi-lo em sílabas poéticas e contá-las conforme as normas.” É preciso contar conforme as normas, ou seja, até a última tônica.

Com essas definições em mente, esta tese considera, portanto, que *escandir* é dividir um verso em sílabas poéticas. O substantivo *escansão* se refere tanto ao processo de escandir quanto ao resultado do processo. De acordo com esta definição, uma dada *escansão* de um verso pode ou não ser correta; para que seja correta, uma *escansão* precisa fazer a divisão das sílabas de acordo com o que é esperado pelo contexto em que o verso aparece. A proposta da tese é justamente produzir *escansões* que sejam tão corretas quanto as feitas por humanos.

A coleção dos ictos de um verso, incluída a sílaba tônica final, chama-se *esquema rítmico* e é comumente denotada pela concatenação dos índices das sílabas acentuadas, separados por um hífen. Assim, o esquema rítmico do seguinte verso é 3-6-10, pois a 3ª, a 6ª e a 10ª sílabas são ictos:

	1	2	3	4	5	6	7	8	9	10	11
GAM.1.1.27	Não	po-	di-	a	mar-	char-	po-	r	um	de-	ser-ito

### 2.1.3. Tipologia

Um verso pode ser classificado, de acordo com a sua métrica e o seu esquema rítmico, em determinados tipos. Alguns recebem nomes explícitos, enquanto outros são descritos por meio do seu metro e esquema rítmico. Os versos menores, embora possuam variação rítmica, não estão tão presos a esquemas pré-definidos quanto os maiores. É a partir dos versos eneassilábicos que os modelos tradicionais possuem mais força.

Os versos de 9 sílabas são bastante comuns nos esquemas 4-9 e 3-6-9. No primeiro caso é usual que eles sejam interpretados como compostos de tetrassilábicos. Alguns exemplos de versos nesse esquema são:

	1	2	3	4	5	6	7	8	9	10
ANJ.151.6.3	Oh!	lu-	a	fri-	a,	vem	me	cho-	rar,	
FIG:2.8.1.3	I-	guai-	s	aos	ven-	to-	s	em	ro-	do-pi- os
SIL:5.50.1.5	Não	ve-	de-	s	an-	jos	de	bran-	cas	ves-tes

O segundo caso é representado pelos seguintes exemplos:

	1	2	3	4	5	6	7	8	9	10		
SIL:5.2.2.1	Co-	mo	é	gra-	to,	nas	lu-	tas	bri-	o-	sas,	
MAG.11.3.1	A	ve-	lhi-	ce,	do	mun-	do	can-	sa-	da,		
MAT:39.9.2.1	Mais	fi-	dal-	go	que	as	mes-	ma-	s	es-	tre-	las,

Muito comuns na poesia em português são os versos decassílabos. O *decassílabo heroico* tem necessariamente a sua 6ª sílaba acentuada, como nos seguintes exemplos:

	1	2	3	4	5	6	7	8	9	10	11
CAM.1.1.1.1	A-	s ar-	ma-	s e os	ba-	rõe-	s a-	ssi-	na-	la-	dos,
MAT:1.1.1.3	Não	sa-	bem	go-	ver-	nar	su-	a	co-	zi-	nha,
DUR.1.1.1.1	De um	va-	rão	em	mil	ca-	so-	s a-	gi-	ta-	dos,
COS.1.3.1	Bem	sei,	que	de ou-	tros	gê-	nio-	s o	Des-	ti-	no,
GON.1.1.1.2	Tão	bem	pin-	ta-	da a his-	tó-	ria	nos	teus	ver-	sos,

O *decassílabo sáfico* tem a 4ª sílaba acentuada e geralmente também a 8ª:

	1	2	3	4	5	6	7	8	9	10	11	
MAG.22.8.2	Do	pó	ti-	ra-	da	pe-	la	mão	do	E-	ter-	no,
PIN.66.42.3	De in-	çar	de	plu-	mas	seus	co-	xins	não	ce-	ssa.	
ANJ.1.3.1	Pai-	ran-	do a-	ci-	ma	dos	mun-	da-	nos	te-	tos,	
FIG:3.1.1.3	No-	ss'al-	ma	can-	ta u-	ma	can-	ção	vi-	bran-	te,	
SIL:1.10.1.2	da	Pri-	ma-	ve-	ra	na es-	ta-	ção	for-	mo-	sa,	

Há versos que podem se encaixar em mais de um padrão, pois nada impede que um verso tenha esquemas rítmicos como 4-6-8-10 ou 2-4-6-8-10. Em casos assim o verso pode ser classificado como híbrido ou então pode-se buscar algum argumento semântico ou sintático para justificar sua classificação como um tipo ou outro. São exemplos de versos ambíguos:

	1	2	3	4	5	6	7	8	9	10	11
CAM.1.5.2	E	não	de a-	gres-	te a-	ve-	na ou	frau-	ta	ru-	da,
DUR.1.8.6	Em	no-	va	li-	ra	can-	to, em	no-	vo	ple-	tro,
GON.1.1.53	Ou	da	fin-	gi-	da	ce-	na os	vão-	s a-	dor-	nos.
PIN.1.19.1	Co-	mo	quem	lu-	cro a-	ne-	la	noi-	te e	di-	a,
FIG:3.2.3.2	De-	scei,	ó	an-	jos,	to-	do o	vi-	nho, e	to-	da

De uso mais restrito é o *decassílabo provençal*, de padrão rítmico básico 4-7-10:

	1	2	3	4	5	6	7	8	9	10	11
MAT:5.25.1.4	as	ma-	ra-	vi,	ma-	ra-	vi,	ma-	ra-	vi-	lhas.
COS.110.55.1	En-	vol-	to o	cor-po	em	um	cân-	di-	do	man-	to,
MAG.45.5.21	Can-	tar	da in-	fân-	cia	tua-	s á-	rias	sau-	do-	sas,
ANJ.15.4.1	Le-	van-	do a-	pe-	nas	na	tum-	ba	car-	ca-	ça
SIL:1.92.1.1	Por	sob	o	véu	d'i-	ma-	cula-	da	bran-	cu-	ra

Este tipo de decassílabo é tão raro que é possível que alguns dos exemplos acima se devam a erros de edição.

Os hendecassílabos possuem comumente o esquema 5-11, como atestam os exemplos abaixo:

	1	2	3	4	5	6	7	8	9	10	11	12
SIL:1.19.4.1	Nas	noi-	tes	se-	re-	nas,	de a-	zul	fir-	ma-	men-	to,
FIG:2.6.1.1	Pe-	lo-	s o-	lhos	tris-	tes	que ho-	je	vi	pa-	ssar	
MAG.42.1.1	Nas	vei-	a-	s o	san-	gue	já	não	me	ga-	lo-	pa,
MAT:39.9.18.2	E	na	ta-	bu-	le-	ta	man-	dou	es-	cul-	pir	

Entretanto, é comum que tais versos sejam na verdade compostos de dois pentassílabos, o que é evidenciado quando se encontra um verso que, dada a regularidade do poema, deveria ser hendecassílabo, mas que contém no máximo 10 ou que precisa de fenômenos forçados para fazer com que as 11 sejam atingidas:

	1	2	3	4	5	6	7	8	9	10	11
FIG:2.22.7.4	E	quem	sa-	be	lá	se,	na	mo-	ci-	da-	de,
MAT:39.9.39.2	Uma	da-	s A-	vós	da	par-	te	vi-	ril,		

Ambos os versos acima encontram-se em contextos em que um hendecassílabo é esperado; a explicação mais plausível é que eles são compostos.

Os dodecassílabos em português quase sempre tem esquema rítmico 6-12, caso em que são chamados de *alexandrinos*, como nos seguintes exemplos:

	1	2	3	4	5	6	7	8	9	10	11	12	13	
ANJ.145.1.1	Au-	ro-	ra	mor-	ta,	fo-	ge!	Eu	bus-	co a	vir-	gem	lou-	ra
SIL:1.1.1.1	Do	sol	no	res-	plen-	dor,	da	lu-	a	na	be-	le-	za,	
FIG:1.1.1.1	Ve-	la-	da-	men-	te o	sol	ves-	tiu	to-	do o	su-	dá-	rio	

Casos de dodecassílabos em que a 6ª não é acentuada existem, mas são raros. Quando o esquema rítmico é 4-8-12, o verso é chamado de *alexandrino trimétrico*, como nos seguintes exemplos:

	1	2	3	4	5	6	7	8	9	10	11	12	13
ANJ.214.1.17	A	ma-	ra-	vi-	lha	das	brí-	lhan-	tes	ma-	ra-	vi-	lhas.
SIL:2.24.3.4	De-	rru-	ba, a-	rras-	ta o	que	lhe	vem	de en-	con-	tro ao	pei-	to.
FIG:1.149.4.3	Bên-	ção	de a-	mor,	bên-	ção	fe-	liz,	bên-	ção	de	vi-	dal

Há poemas ou estrofes em que aparecem versos de duas medidas:

uma maior, tipicamente de 9 ou mais sílabas, e outra menor, geralmente de 6 ou menos. A medida menor frequentemente tem um tamanho correspondente a uma posição silábica importante para o verso maior. Os quebrados de decassílabos, por exemplo, são bastante comuns; os seguintes versos maiores e menores aparecem em sequência nos poemas de onde foram tirados:

	1	2	3	4	5	6	7	8	9	10	11	
MAG.1.1.1	Quan-	do	da	noi-	te	o	véu-	ca-	li-	gi-	no-	so
MAG.1.1.2	Do	mun-	do	me	se-	pa-	ra,					
	1	2	3	4	5	6	7	8	9	10	11	
ANJ.25.37.3	Viu-	to-	da	a	po-	dri-	dão	de	su-	a	ra-	ça
ANJ.25.37.4	Na	tum-	ba	de	l-	ra-	ce-	mal...				

Similarmente encontram-se quebrados de alexandrinos, eneassílabos e hendecassílabos:

	1	2	3	4	5	6	7	8	9	10	11	12	13					
FIG:2.2.15.1	Vi-	nha	ao-	le-	me,	o	Jo-	ão;	e	o	Pe-	dro,	vi-	nha	à	es-	co-	ta
FIG:2.2.15.2	De-	ve-	la	to-	da	ro-	ta...											
	1	2	3	4	5	6	7	8	9	10								
ANJ.151.3.1	Bai-	xa,	por-	tan-	to,	co-	ra-	ção	er-	mo								
ANJ.151.3.2	De	lu-	a	fri-	a													
	1	2	3	4	5	6	7	8	9	10	11	12						
MAG.42.1.3	A	li-	ra	ca-	no-	ra	do	cis-	ne	Be-	ó-	cio						
MAG.42.1.4	Dei-	xei-	so-	bre	a	trí-	po-	de.										

#### 2.1.4. Metaplasmos

Quando um verso é corretamente escandido, frequentemente certos fenômenos, chamados metaplasmos, são observados. A pronúncia em uma sílaba de duas vogais que normalmente ficam em duas, por exemplo, é comum na poesia. Os manuais de versificação com frequência descrevem estes fenômenos do ponto de vista do poeta. Tavares (166), por exemplo, exemplifica a paragoge, a adição de um fonema ao fim da palavra, com as palavras *fugace* e *faze*; Chociay (75), similarmente, descreve o emprego de *Mavorte* em vez de *Marte* como uma ampliação. Esta tese busca exercer o papel de leitor de poesia, não de poeta, e portanto os metaplasmos são descritos somente do ponto de vista do leitor.

**Sinalefa.** Quando duas palavras são postas em contato no interior do verso, as vogais que ficam expostas no fim de uma e no início de outra interagem. Quando os encontros vocálicos são preservados fala-se de *dialefa* ou *hiato*; quando as vogais se misturam ou se perdem fala-se de *sinalefa*. Em alguns casos, a pronúncia naturalmente mantém as vogais em sílabas

distintas, enquanto em outros é natural que elas se juntem. O nome *dialefa*, portanto, pode significar tanto a separação de vogais que normalmente se juntariam quanto a manutenção de um hiato natural. Nos seguintes versos, por exemplo, o primeiro contém uma dialefa de separação entre *e* e *a*, enquanto o segundo, uma de manutenção entre *até* e *ao*:

1    2    3    4    5    6    7    8    9    10    11  
MAG.4.8.9 En-tre-Deu-s, e a-te-rra-co-lo-ca-do,  
ANJ.12.18.3 E a-té ao fim, cor-ta-sse e re-cor-ta-sse

O mesmo raciocínio aplica-se à sinalefa. O primeiro exemplo abaixo mostra uma sinalefa mais forçada enquanto o segundo uma mais natural:

1    2    3    4    5    6    7    8    9    10    11  
ANJ.55.10.2 — Es-ta- rei mor-to?! E a es-ta per-gun-ta es-tra-nha  
FIG:5.1.1.4 Por so-bre os cam-po- s e por so-bre os ri-os.

No primeiro caso, o poeta solicita que o leitor aglutine numa só sílaba a vogal final de *morto*, a conjunção *E*, o artigo *a* e a vogal inicial de *esta*. No segundo, a preposição *sobre* é naturalmente unida ao artigo que segue.

Dois casos particulares de sinalefa merecem destaque. A *crase* é a união de vogais similares, como nos exemplos abaixo, em que as vogais escritas por *o* e *a*, respectivamente, são fundidas:

1    2    3    4    5    6    7    8    9    10    11  
PIN.1.4.2 Tan-to o so-no os sen-ti-dos me-to-ma-ra,  
CAM.1.18.8 E cos-tu-mai-vos já a se-r in-vo-ca-do.

A *elisão* remove a vogal final de uma palavra, tipicamente um *a* na escrita, quando a palavra seguinte começa com vogal, como nos exemplos a seguir:

1    2    3    4    5    6    7    8    9    10    11  
DUR.1.1.8 Pois só co-nhe-ço he-rói-quem ne-la é for-te.  
GON.1.1.11 Da-rai-va e do pra-zer. Ma-s ah! que di-sse!

**Sinérese.** A *sinérese* opera de forma similar à sinalefa, mas no interior dos vocábulos: trata-se da pronúncia numa sílaba de vogais que normalmente ocupariam duas. A *diérese* é o oposto, a separação em duas sílabas de sons que normalmente seriam pronunciados numa sílaba só. Também como no caso da sinalefa, a sinérese pode ser a efetiva junção ou a manutenção da junção natural, e a diérese, a preservação de uma separação que tipicamente ocorre ou a separação forçada.

São exemplos de sinéreses a pronúncia das palavras *piadosa* e *coroada* nos versos abaixo:



	1	2	3	4	5	6	7	8	9	10	11
PIN.16.16.3	En-	tre e-	les,	da al-	ma na	ex-	pan-	são	pie-	do-	sa;
ANJ.261.4.2	Deu-	sa	co-	roa-	da,	sem	que al-	guém	pin-	tar-	te

Para a identificação da sinérese, pouco importa se a pronúncia mais natural destas palavras é em 3 ou 4 sílabas; o importante é que foram pronunciadas com 3.

Já as palavras *circuitos* e *sai* sofrem diérese nos versos abaixo:

	1	2	3	4	5	6	7	8	9	10	11
PIN.17.33.2	Em	cir-	cu-	i-	tos	lar-	gos	sem	ter	pre-	ssa;
CAM.10.137.5	On-	de	sa-	i	do	chei-	ro	mais	per-	fei-	to

**Aférese, síncope, apócope.** Estes três metaplasmos correspondem à remoção de uma vogal em três posições diferentes: *aférese* no início da palavra; *síncope* no seu interior; *apócope* no fim. A aférese é particularmente comum quando a palavra começa com *es* mais uma oclusiva, como nos exemplos a seguir, nas palavras *espaço* e *Estuda*:

	1	2	3	4	5	6	7	8	9	10	11	
PIN.31.11.3	E al-	çam	bus-	to-	s em	tor-	no ao	es-	pa-	ço hi-	an-	te".
SIL:5.189.3.4	"Estu-	da a	li-	ção!"								

A aférese com frequência é indicada no próprio texto, seja através da omissão da vogal ou de um apóstrofo, poupando ao leitor a inconveniência de chegar ao fim do verso para descobrir que escandiu errado:

	1	2	3	4	5	6	7	8	9	10	11
PIN.5.7.1	"O-	lha	co-	mo en-	tra-	s e em	quem	stás	fi-	an-	do:
MAG.11.4.4	Stá	só	na	mu-	dan-	ça;					

A síncope ocorre em várias classes de palavra, tipicamente removendo uma vogal átona entre duas consoantes:

	1	2	3	4	5	6	7	8	9	10	11
SIL:1.105.4.1	a	se a-	gru-	pa-	r em	cú-	mulos	no ho-	ri-	zon-	te,
FIG:5.31.1.4	Do	di-	vi-	no	po-	mar	da	feli-	ci-	da-	de...
MAG.44.3.3	Do	tu-	rí-	bulu	de	mi-	nha al-	ma			
SIL:1.121.2.3	na	bran-	ca	pé-	tala	da	flor;				

Similarmente à aférese, a síncope é, no mais das vezes, indicada no texto original. Os apóstrofes nos próximos versos indicam síncoptes marcadas na edição:

	1	2	3	4	5	6	7	8	9	10	11
PIN.44.12.2	Tan-	tos	sé-	c'los	com	lá-	gri-	mas	pe-	di-	do,
ANJ.169.1.4	Fri-	a	co-	mo um	cre-	pús-	c'lo	da	Ju-	dei-	a.
MAG.29.7.10	De áu-	reas	c'ro-	as,	de	púr-	pu-	ra	co-	ber-	tos,

A apócope é mais rara que a aférese e a síncope. De exemplo há o

seguinte verso, em que palavra *mármore* perdeu sua vogal final:

MAT:38.21.1.46 des-|te|már-|more,|des-|te|ro-|ble,

Também a apócope costuma ser marcada no texto original, como nos exemplos abaixo:

SIL:1.1.8.3 no|val'|do|pran-|to e|dor|nã|dei-|xes|va-|ci-|lar  
DIA.2.24.3 A|pel'|vis-|to-|sa;|sob|a-|s hir-|tas|cer-|das,

**Epêntese, paragoge.** Em certas circunstâncias uma vogal de apoio é inserida após uma consoante. Quando esta inserção é feita no interior da palavra, recebe o nome de *epêntese*; ao fim chama-se *paragoge*. Estes metaplasmos nunca vem marcados no texto original, devido ao desfiguramento que causariam às palavras.

A epêntese visa sobretudo a desfazer grupos consonantais complexos, como nos seguintes exemplos, em que os grupos consonantais *gn*, *ct* e *bs* devem ter um som átono de *i* após seu primeiro elemento:

MAG.28.15.5 Oh|be-|ni-|gn-|o|sol,|que a|vi-|da a-|que-|ces,  
SIL:5.184.4.3 Sa-|ú-|da o|for-|te in-|vi-|ct-|o|Bra-|sil.  
SIL:5.61.1.7 ja-|mai-|s o-|bs-|cu-|re-|ci-|das|

A paragoge opera sobretudo na palavra *sob*, adicionando-lhe um *i* átono final tão comum em sua pronúncia corriqueira:

FIG:4.51.1.3 So-|b noi-|tes tris-|tí-|ssi-|mas,|pre-|ssa-|gas,  
PIN.27.15.3 So-|b ver-|de le-|ão,|so-|fre mes-|qui-|nha.  
SIL:1.96.1.6 So-|b fu-|né-|reo ci-|pres-|te|

**Sístole, diástole.** Há casos em que o poeta requer que uma palavra tenha a posição de seu acento modificada. Quando o acento recua mais para o início, o nome do fenômeno é *sístole*; quando ele avança mais para o fim, *diástole*. Às vezes é difícil dizer se a pronúncia da sístole, que é muito mais comum que a diástole, era característica da pronúncia do poeta ou se foi de fato um artifício empregado conscientemente. De qualquer modo, a sístole às vezes vem marcada no texto original com uma mudança ortográfica:

ANJ.25.104.4 Pa-|ra al-|can-|çar|de-|poi-|s a|pe-|ri-|fé-|ria!  
CAM.1.19.8 Que-|do|ga-|do-|de|Pró-|teo|são|cor-|ta-|das.

Há casos também em que a sístole não está marcada no original e fica a critério do leitor fazê-la:

	1	2	3	4	5	6	7	8	9	10	11
CAM.10.103.3	Que	se in-	ju-	ria	de u-	sar	fun-	di-	do	co-	bre
PIN.56.46.3	So-	bre as	fo-	lha-	s em	ro-	cio as	go-	tas	flu-	em.
SIL:1.57.4.5	nem h-	ar-	mo-	nia	mais	ra-	ra				

Nos versos acima, para cada verso respectivamente, as pronúncias *injúria*, *rócio* e *harmônia* devem ser adotadas.

A diástole, mais rara, também normalmente vem refletida na ortografia, como no verso seguinte de Camões, em que *Etíope* torna-se *Etiópe*, produzindo um verso sáfico:

	1	2	3	4	5	6	7	8	9	10	11
CAM.5.62.2	*Pos-	to	que	to-	do-	s E-	ti-	o-	pe-	s e-	ram,

Como exemplo de diástole não marcada há o seguinte, em que também poderia ter sido feita a sístole:

	1	2	3	4	5	6	7	8	9	10	11
FIG:4.39.2.1	E	te-	ria-	mos	go-	za-	do o	que	go-	za-	mos

A pronúncia *teríamos* permite que o verso seja escandido nas 10 exigidas pelo metro.

**Ectlipse.** A *ectlipse* consiste no apagamento da trava nasal final de certas palavras, permitindo-lhes entrar em sinalefa. Novamente trata-se de um fenômeno que pode ou não vir marcado com apóstrofo no texto original. Exemplos de casos marcados são:

	1	2	3	4	5	6	7	8	9	10	11	12
MAG.42.3.3	Dos	no-	ssos	pra-	ze-	re-	s o-	ri-	ge'i-	ne-	fá-	vel;
PIN.72.10.1	"Por	con-	tra-	to, en-	tre	Deu-	s e o h-	o-	me'a-	cei-	to,	
SIL:5.178.2.1	Que	Vir-	ge' é	e-	ssa	de	be-	le-	za	tan-	ta,	

O caso mais comum de ectlipse é observado com a preposição *com*; os seguintes exemplos não estão marcados no texto:

	1	2	3	4	5	6	7	8	9	10	11
PIN.22.3.3	Com ar-	tes	no-	va-	s ou	en-	tre	nó-	s u-	sa-	das:
ANJ.36.16.3	Com a	fle-	xi-	bi-	li-	da-	de	de um	mo-	lus-	co,
SIL:1.59.1.1	Quem,	com a	plu-	ma-	gem	do	gua-	rá	for-	mo-	so,

Há também exemplos mais complexos, como os abaixo, em que a ectlipse ocorre nas palavras *soluçam*, *não* e *tingem*:

	1	2	3	4	5	6	7	8	9	10	11	12	13
FIG:1.129.1.4	On-	de em	má-	goas	so-	lu-	çam o	ven-	to	sul'e o	nor-	ite.	
CAM.3.29.2	Que em	tan-	ta an-	ti-	gui-	da-	de	não há	cer-	te-	za,		
SIL:5.194.3.3	Que	tin-	gem as	nu-	vens	do	céu,						

Tabela 2.2.: Classificação das estrofes segundo número de versos.

Versos	Nome	Alternativa
2	Dístico	Parelha
3	Terceto	
4	Quarteto	Quadra
5	Quinteto	Quintilha
6	Sexteto	Sextilha
7	Sétima	Setilha
8	Oitava	
9	Nona	Novena
10	Décima	

## 2.2. A estrofe

As *estrofes* são agrupamentos de versos, separados graficamente uns dos outros por espaço. As estrofes podem ser classificadas com relação ao número de versos que elas contêm, como descreve a Tabela 2.2. A utilidade de uma tal classificação consiste em poder se referir de forma breve e específica a estrofes de determinado número de versos; é compacto dizer que um soneto é feito de dois quartetos e dois tercetos. A classificação das estrofes de acordo com seu número de versos acaba se misturando à sua classificação de acordo também com as rimas e o número de sílabas dos versos. Chociay (75) e Tavares (166) observam, por exemplo, que tradicionalmente *quintilha* se referiria às estrofes de cinco versos com sete sílabas ao menos, o termo *quinteto* estando reservado às demais estrofes de cinco versos. Nesta tese os termos são adotados sem nenhuma pretensão de caracterização da estrofe para além do número de versos.

A existência de estrofes de um verso, que se chamariam *monósticas*, não é muito bem estabelecida. Goldstein (99) e Chociay (75) nem mesmo as incluem em suas enumerações, argumentando que os versos que aparecem graficamente isolados no texto pertencem, organicamente, a outro (75). Mattoso (126) nomeia estrofes ainda maiores, como as *undécimas* e *duodécimas*, enquanto outros as chamam de irregulares (166) ou apenas comentam que são raras (75).

As estrofes podem ser classificadas segundo a regularidade do número de sílabas que seus versos possuem. Uma estrofe em que todos os versos possuem a mesma quantidade de sílabas chama-se *isométrica*; quando esta quantidade não é constante, a estrofe recebe o nome de *heterométrica*. Estrofes em que a quantidade de sílabas dos versos varia bastante são cha-

madras *polimétricas*.

### 2.3. O poema

Um *poema*, para os propósitos desta tese, é um agrupamento de estrofes. Dois versos próximos um ao outro, dentro da mesma estrofe ou não, *ri-mam* se eles soam parecidos a partir da última vogal tônica. Esta tese não investiga nenhum aspecto ligado à rima, de maneira que a terminologia a ela associada não será aqui definida.

Os mesmos atributos métricos aplicados à estrofe podem também ser ditos do poema: um poema em que os versos todos tem o mesmo comprimento é *isométrico*; caso contrário é *heterométrico*. Quando há um grande número de metros, o poema é *polimétrico*.

Segundo o tamanho das estrofes, bem como outros atributos da rima, da métrica e do ritmo, poemas recebem nomes específicos. Em particular, um *soneto* é um poema isométrico composto de dois quartetos e dois tercetos. Em geral é composto de decassílabos ou dodecassílabos. O esquema das rimas dos sonetos é variável, mas a existência das rimas é um atributo bem estabelecido.



### 3. TRABALHOS RELACIONADOS

1	2	3	4	5	6	7	8	9	10	11
Eu	que-	ro	com-	po-	r um	so-	ne-	to	du-	ro
co-	mo	poe-	ta al-	gum	ou-	sa-	ra es-	cre-	ver.	
Eu	que-	ro	pin-	ta-	r um	so-	ne-	to es-	cu-	ro,
se-	co, a-	ba-	fa-	do,	di-	fi-	cil	de	ler.	

Drummond

Este capítulo discute as publicações científicas que, aplicadas a qualquer língua, buscaram processar poesia do ponto de vista métrico ou rítmico com base nos versos em sua forma textual. Em particular, não são descritos artigos que tratam de rima, que são vários (158, 141, 76). Uma recente *survey* foi feita sobre a análise automática de poesia (80); as referências lá contidas serviram para ampliar o capítulo mas ela em si não é aqui discutida.

Foram definidos cinco graus de relevância para os idiomas com trabalhos encontrados, conforme o nível de relação genética e cultural com o português. O espaço foi alocado aos trabalhos relacionados conforme seu grau de relevância e o detalhamento presente no artigo. São os graus de relevância:

- Grau 0. A própria língua portuguesa, cujas propostas são descritas amplamente.
- Grau 1. As línguas românicas: espanhol, italiano, francês e provençal. Estas línguas recebem bastante atenção nesta revisão, com uma seção dedicada a cada língua e uma subseção a cada proposta.
- Grau 2. As línguas europeias de modo geral: inglês, alemão, russo e tcheco. Menos espaço é dedicado a estas línguas, cada uma recebendo uma subseção. Apenas alguns trabalhos destacados são descritos em algum detalhe; os demais tem apenas uma ou duas principais características mencionadas.
- Grau 3. As línguas clássicas: grego antigo e latim. Todos os trabalhos são descritos de forma breve numa subseção conjunta.

- Grau 4. As demais línguas: sânscrito (e parentes), árabe e turco. Os trabalhos encontrados são citados de forma condensada sem direito a subseção.

A seguir são, portanto, seguem os trabalhos relacionados. Ao fim do capítulo são apresentadas ainda algumas considerações finais.

### 3.1. Português

Existiu apenas um esforço de pesquisa que contemplasse a escansão de versos em português; trata-se do trabalho orientado por Nuno Mamede, de Portugal, que produziu, entre outros, duas dissertações de mestrado (40, 124). Este esforço será examinado na subseção abaixo.

Há ainda uma tese de doutorado (161) com o provocativo título *Como Construir um Leitor de Poesia*. O trabalho, apesar de estar longinquamente relacionado a programas de leitura de poesia, citando inclusive as produções de Mamede, é bastante teórico e situado na área de Teoria da Literatura, pouco ou nada dizendo de relevante à presente tese, que tem preocupações práticas.

Recentemente, Carvalho (71) iniciou um projeto para a detecção de padrões poéticos em prosa na língua portuguesa. Resultados mais concretos, porém, ainda não foram publicados.

#### 3.1.1. LuCas e SAEP

Estes dois projetos foram orientados por Nuno Mamede. As publicações podem ser divididas em dois estágios: a pesquisa inicial, realizada por Paulo Araújo, descrita num artigo (39) e numa dissertação (40); o estágio final, representado por dois artigos muito similares em inglês (123, 122) e uma dissertação escrita por João Marques (124). De modo geral, o estágio inicial gerou o programa LuCas, chamado assim em homenagem a Luís de Camões, que é capaz de contar sílabas porém não de resolver metaplasmos; o estágio final produziu o SAEP, o Sistema de Apoio à Escrita de Poemas, com habilidades limitadas mas existentes de lidar com os metaplasmos.

Ambos os sistemas compartilham diversas características. Eles recebem como entrada um poema codificado em texto simples e dividem-no em versos e estrofes de acordo com as linhas vazias. As palavras dos versos são transcritas foneticamente através de um módulo externo chamado DIXI (139), que também faz a divisão silábica gramatical delas. Os sistemas têm o objetivo tanto de analisar e classificar poemas, estrofes e versos



Tabela 3.1.: Contagem de sílabas pelos sistemas LuCas e SAEP dos versos de uma mesma estrofe de Camões.

Verso	LuCas	SAEP
As armas e os barões assinalados	12	10
Que da ocidental praia lusitana	12	10
Por mares nunca de antes navegados	12	10
Passaram ainda além da Taprobana,	13	10
Em perigos e guerras esforçados	12	10
Mais do que prometia a força humana,	13	10
E entre gente remota edificaram	13	10
Novo Reino, que tanto sublimaram;	11	10

quanto de auxiliar o poeta em sua arte. Ambos dispõem, por exemplo, da funcionalidade de sugestão da palavra final de um verso inacabado, que é obtida considerando-se informações métricas e rítmicas derivadas de um verso de referência.

Nenhuma menção é feita com relação ao tratamento da pontuação, de estrangeirismos ou de palavras não escritas por extenso no texto de entrada. Apesar de realizarem análise sintática através de um módulo externo, os resultados desta análise não são utilizados na escansão dos versos.

O sistema LuCas tem poucas habilidades no que se refere à análise métrica dos versos, como evidencia a Tabela 3.1. Ele não detecta nenhum metaplasmo, concatenando as transcrições fonéticas das palavras umas às outras. Ademais, ele conta todas as sílabas do verso, incluindo aquelas além da última tônica. Assim, o sistema LuCas conta 13 sílabas no seguinte verso de Camões:

1 2 3 4 5 6 7 8 9 10 11 12 13  
Pas-sa-ram|a-in-da|a-lém|da-Ta-pro-ba-na,

A divisão silábica acima é apenas presumida, pois não consta nas publicações.

Já o SAEP introduz alguns mecanismos para a resolução de encontros vocálicos. Ele aplica regras à transcrição fonética para que a divisão silábica seja ajustada. Assim, como mostra a Tabela 3.1, o novo sistema corretamente contou o número de sílabas na estrofe inteira de Camões. As regras são aplicadas pelo sistema, contudo, de forma compulsória, o que é problemático pois os metaplasmos no português são raramente ou nunca obrigatórios. Em seus experimentos, o autor analisou outra estrofe de Camões, em que o sistema contou incorretamente o número de sílabas de 5

Tabela 3.2.: Contagem de sílabas pelo sistema SAEP dos versos de outra estrofe de Camões.

Verso	SAEP
Ó que famintos beijos na floresta,	10
E que mimoso choro que soava!	9
Que afagos tão suaves, que ira honesta,	9
Que em risinhos alegres se tornava!	10
O que mais passam na manhã, e na sesta,	11
Que Vénus com prazeres inflamava,	10
Melhor é experimentá-lo que julgá-lo,	11
Mas julgue-o quem não pode experimentá-lo.	12

dos 8 versos, como mostra a Tabela 3.2. Os dois últimos versos da tabela apresentam um número a mais de sílabas do que teriam se uma outra edição fosse utilizada nos experimentos; as duas ocorrências de *experimentá-lo*, na edição de referência adotada pela presente tese para o *corpus* que coincidentemente inclui a estrofe da tabela, estão grafadas *expriméntá-lo*, o que facilita a escansão do verso. Mesmo com essa alteração, 4 dos 8 versos continuariam errados.

Os versos incorretamente escandidos da Tabela 3.2 são os seguintes, com divisão silábica relatada pelos autores:

1	2	3	4	5	6	7	8	9	10	11	12	13
E	que	mi-	mo-	so	cho-	ro	que	soa-	va			
Que a-	fa-	gos	tão	sua-	ves	que i-	ra ho-	nes-	ta			
O	que	mais	pa-	ssam	na	ma-	nhã	e	na	ses	ta	
Me	lhor	é ex-	pe-	ri-	men-	tá-	lo	que	jul-	gá-	lo	
Mas	jul-	gue-	o	quem	não	po-	de ex-	pe-	ri-	men-	tá-	lo

Os sistemas LuCas e SAEP estão primariamente preocupados com a contagem do número de sílabas dos versos, e não com a escansão. A discussão dos autores sobre os problemas encontrados concentra-se, portanto, na questão numérica. Como trabalhos futuros, por exemplo, cita-se (124) a apresentação, na interface do programa, do número mínimo e máximo de sílabas que o verso pode conter, o primeiro sendo computado com todas as sinalefas e sinéreses, o último com as dialefas e diéreses. Comenta ainda o autor que, ao serem considerados também os demais versos do poema, o programa poderia apresentar um número provável de sílabas de cada verso. Tal abordagem é similar àquela adotada pelo Aoidos, que determina o metro de um verso com base no contexto em que ele se en-

contra. Contar o número de sílabas poéticas de um verso é suficiente para atribuir rótulos como *heptassílabo* e *decassílabo*, mas a escansão depende, além deste número, da correta distribuição das sílabas no interior do verso. Para escandir um verso, não basta conhecer seu número de sílabas; é preciso saber quais metaplasmos devem ser realizados e quais não.

O LuCas foi testado com aproximadamente duzentas estrofes compostas por crianças — o que evidencia o foco bastante diferente deste sistema em comparação com o Aoidos —, dez quadras de António Aleixo e duas estrofes dos *Lusiadas* de Camões. Os testes, todavia, avaliaram apenas o tempo de resposta do sistema, e não sua acurácia. Já o SAEP teve sua acurácia estimada a partir de um conjunto de 12 poemas que totalizam 197 versos. Em todos os critérios de avaliação, como classificação das rimas e das estrofes, o sistema não obteve nenhum erro — exceto justamente na contagem de sílabas poéticas, em que o sistema errou em 35 versos, o que resulta numa taxa de acerto de 82,2%. Este número de erros, porém, leva em consideração poemas com versos livres, que tipicamente não são objeto de análise das ferramentas de escansão automática. Se reduzirmos o campo de visão aos 105 versos dos 7 poemas sem versos livres ou polimétricos, o número de erros é 24, o que resulta numa taxa de acerto de 77,1%. A avaliação, entretanto, não verificou a escansão dos versos, apenas o número de sílabas poéticas.

Do ponto de vista da escansão, a principal limitação dos sistemas é a aplicação compulsória dos metaplasmos e a preocupação primária com o número de sílabas e não com a correta distribuição delas no verso. Com os sistemas LuCas e SAEP, os autores pretendem incentivar a produção e o estudo da poesia em português, inclusive no contexto escolar.

### 3.2. Espanhol

A poesia em espanhol, assim como a em português, resolve encontros vocálicos intervocabulares de forma não-determinística: o mesmo encontro vocálico pode ser resolvido de formas distintas, conforme o contexto. Os desafios encontrados na escansão do espanhol são, portanto, parecidos com aqueles do português.

Similarmente ao português e diferentemente do italiano, a posição do acento, tão fundamental para o poesia, pode ser localizada de forma sistemática. O espanhol possui, com efeito, uma ortografia de base bastante fonética, o que significa que os trabalhos da área com frequência não precisam fazer uma transcrição fonética; o texto original já está em formato

suficientemente adequado.

As subseções seguintes descrevem as três propostas encontradas na literatura: o Colors of Poetry, o programa em Prolog de Gervás, e a ferramenta de Navarro-Colorado.

### 3.2.1. Colors of Poetry

Robinson (160) apresentou um programa chamado Colors of Poetry que analisa poemas em espanhol, exibe-os de forma gráfica e auxilia poetas na busca por palavras. A inspiração do autor foi a existência de ferramentas feitas para artistas que trabalham em outros meios, como a imagem e o som. O autor produziu uma dissertação de mestrado com um foco bastante grande na apresentação de sua ferramenta a um público da área de artes e literatura, descrevendo em poucos detalhes, e apenas em anexo, o funcionamento do programa. Nenhum artigo publicado pelo autor foi encontrado.

O documento não descreve vários detalhes de interesse à presente tese. Ele não contém informações sobre o formato de entrada e possíveis etapas de pré-processamento. O autor não descreve como o acento é encontrado, limitando-se à explicação de que o programa segue as regras gramaticais do espanhol. A ferramenta realiza a separação silábica, mas o trabalho não especifica como a faz.

O Colors of Poetry usa um esquema bastante simples para a transcrição fonética. O autor descreve-o em termos dos fonemas que não possuem uma correspondência biunívoca com os grafemas do espanhol: o /k/ de *po[ll]o*, o /r/ de *ca[rr]o* e *[r]equete*, o /θ/ de *[c]inco* e *[z]orro* (no caso do castelhano), o /tʃ/ de *[ch]oza*, o /k/ de *[c]asa* e *[qu]ebradizo*, o /x/ de *[j]abón* e *[g]ente* e o /b/ de *[v]aya* e *[b]onito*.

O sistema não tem a intenção de produzir uma escansão definitiva. A partir das etapas anteriores, os versos do poema são exibidos com a ortografia original e separação silábica fonética. A transcrição fonética faz algumas sinalefas que o sistema considera obrigatórias, como em *[ye]n* e *penumbr[a i]uminada*; outros casos de sinalefas, sinéreses e diéreses são indicados graficamente, mas o programa não busca estabelecer se elas existem de fato. A este respeito, o autor argumenta que este é o caminho mais seguro e limitado: a alternativa seria comparar os versos com modelos fixos e uns com os outros para definir qual padrão os versos devem seguir e, finalmente, definir quais sinalefas ou diéreses precisam ser feitas para se chegar à métrica correta. O verso a seguir foi analisado pelo Colors of Poetry; das quatro sinalefas encontradas pelo sistema, apenas três

devem ser realizados para que as 10 sílabas métricas necessárias sejam obtidas:

1 2 3 4 5 6 7 8 9 10  
Y|tú, a-|rri-|ba, en|lo al|to, |ver-|de y|frí-|o,

O Colors of Poetry não faz uma análise rítmica automática. Em vez disso, exhibe graficamente vários parâmetros, como o acento tônico e possíveis pausas, que ajudam o leitor a encontrar o ritmo no poema. Por não resolver fenômenos como as sinalefas, o sistema não pode dispor as sílabas numa grade métrica.

A pontuação não é ignorada pelo sistema. Em dois casos ela é considerada: primeiro, ela é utilizada na marcação das possíveis sinalefas, mas o trabalho não descreve exatamente em quais circunstâncias. A julgar pela análise de um poema produzida pelo sistema, certos sinais de pontuação, como ! e j, impedem a marcação de sinalefa, mesmo que a análise manual do mesmo poema exija que a sinalefa cruze os dois sinais de pontuação.

A pontuação também é usada para marcar visualmente as pausas dentro do poema: um verso que termine sem pontuação ganha à sua direita uma fina barra vertical azul, indicando a pausa mínima; um verso que termine com pontuação ganha barras duplas, indicando uma pausa média; o último verso da estrofe é marcado por um bloco sólido de mesma cor. As pontuações que aparecem dentro dos versos também ganham uma barra vertical.

À exceção de algumas comparações descritivas, nenhuma tentativa sistemática de avaliação foi feita de qualquer parte do sistema. O autor aponta como uma das principais limitações o fato de o programa levar em consideração apenas o acento léxico, não se propondo a encontrar o acento rítmico do verso. Por não utilizar nenhuma análise sintática, o Colors of Poetry também não pode distinguir entre homógrafos cuja pronúncia depende da sintaxe, como *para*, que, a exemplo do português, pode ser tanto a preposição átona quanto a 3ª do singular do presente indicativo do verbo *parar*.

### 3.2.2. O sistema de Gervás

Gervás (98) descreveu uma ferramenta que realiza a escansão de versos em espanhol; ele não nomeou sua criação. A partir dos versos em seu formato original, o sistema divide as palavras em sílabas, determina a sílaba tônica de cada palavra, realiza as sinalefas entre as palavras e extrai as rimas de cada verso. É notável que o autor tenha escrito o programa em Prolog, uma linguagem que intuitivamente parece ser muito adequada à tarefa.

Duas aplicações são previstas no artigo: a ferramenta pode ser útil para estudos literários da poesia em espanhol e pode também ser aplicável num contexto pedagógico, tanto na escola quanto na universidade, por mostrar aos alunos como a poesia em espanhol pode ser analisada.

O programa começa pela divisão silábica das palavras, sem fazer transcrição fonética. O sistema cria uma série de grupos para cada palavra: alguns contêm certas combinações de grafemas, como os dígrafos *ch* e *rr*, enquanto outros contêm apenas um. A cada grupo é atribuída uma de 14 possíveis categorias, como *consoante dupla* e *vogal alta*. Tais propriedades são utilizadas não apenas na divisão silábica, mas também, mais tarde, na identificação das sinalefas. Em seguida ocorre um reagrupamento em que ficam juntos os grafemas consonânticos e os grafemas vocálicos que não formam hiato, o que equivale a dizer que este agrupamento de grafemas vocálicos faz a distinção entre diérese e sínese. Este estágio leva em consideração o diacrítico agudo do espanhol: a combinação *aí* forma um hiato, mas *ái* não. A separação silábica final é obtida tomando como base as vogais identificadas, ao redor das quais são distribuídos os grupos consonantais.

Além da utilização do agudo como marcador de tonicidade, as regras para a obtenção da sílaba tônica não são discutidas, mas elas são “poucas e simples”.

Para que a escansão do verso seja obtida, falta apenas que as sinalefas sejam resolvidas. Para atingir este objetivo, o sistema verifica a fronteira entre duas sílabas adjacentes quaisquer: se o fim da primeira e o início da segunda tiverem as vogais apropriadas, como em *te a*, elas são reunidas numa sílaba só. Esta verificação é recursiva: a sílaba resultante de uma sinalefa pode ela mesma sofrer sinalefa. Os tritongos são produzidos justamente pela aplicação dupla da regra de sinalefa. É importante salientar que a sinalefa é realizada de forma obrigatória pelo sistema, ou seja, os mesmos encontros vocálicos são resolvidos sempre da mesma forma.

Eis um exemplo de escansão produzida pelo sistema de Gervás, em que as sílabas em sublinhado são aquelas identificadas como tônicas:

1	2	3	4	5	6	7	8	9	10	11
Ce-	rrar	po-	drá	mis	o-	jos	la	pos-	tre-	ra

O sistema é capaz de identificar as sílabas tônicas lexicais dos versos, mas não busca identificar o acento rítmico. Tal identificação poderia, por exemplo, guiar a aplicação das sinalefas para que as tônicas ficassem na posição necessária. O artigo de Gervás chega a identificar tipos de versos segundo a distribuição das tônicas no verso, mas em sua abordagem os

modelos de versos são utilizados somente na classificação final, quando o processo como um todo já foi terminado.

A avaliação do sistema foi feita com 64 sonetos, todos decassílabos, de seis autores do chamado século de ouro espanhol (aproximadamente entre 1492 e 1659), totalizando 896 versos. Após processar esse *corpus* com o sistema, foram manualmente verificados todos os versos aos quais o sistema atribuiu um número de sílabas métricas diferente de 10. Não está claro se podem existir erros entre os versos não verificados manualmente, mas pode-se conjecturar que o sistema pode, no mesmo verso, ter feito uma sinalefa desnecessária e não ter feito outra necessária, acertando o número de sílabas mas errando na disposição delas. A taxa global de acertos do sistema foi de 88,7% dos versos, variando entre os autores de 81,3 a 95,2%, o que talvez indique que as regras de sinalefas e sinéreses do sistema funcionem melhor para uns autores do que para outros.

O autor verificou que, entre os versos erroneamente classificados pelo sistema, havia duas fontes principais de erro. A primeira seria a licença poética dos autores, que permite que o mesmo verso tenha um número maior ou menor de sílabas conforme seu contexto. A segunda é o tratamento da conjunção *y*, que os poetas ora usam de forma vocálica, ora o fazem semivocalicamente. De fato, se a primeira fonte é desconsiderada, a taxa de acerto sobe para 94,0%; se ambas o são, ela sobe para 99,3%.

O sistema não faz nenhum tipo de processamento especial nos estrangeirismos. Ele possui, contudo, um banco de dados de palavras pré-processadas, que serve um objetivo duplo: tanto acelera o processamento do sistema, servindo efetivamente de *cache*, quanto permite que o usuário adicione palavras, incluídos aí os estrangeirismos, e especifique particularidades como a separação silábica e acento tônico. O artigo de Gervás não faz comentários quanto à pontuação, porém ela é presumivelmente descartada, dada a ausência de qualquer descrição com relação ao seu tratamento e a inexistência de pontuação em exemplos.

Gervás aponta que as sinalefas não estão sendo realizadas a contento. A isto deve-se adicionar que, como ele mesmo afirma, o mesmo verso pode ter contagens silábicas diferentes em contextos diferentes; para resolver de forma mais definitiva as sinalefas é preciso levar em consideração justamente o contexto.

### 3.2.3. O sistema de Navarro-Colorado

Navarro-Colorado (135, 137, 136) propôs uma ferramenta que anota metricamente sonetos em espanhol e produz análises semânticas com base

nas palavras dos versos. Diferentemente das outras duas abordagens para o espanhol, esta ataca o problema dos versos que inicialmente parecem ter mais do que as sílabas que a métrica exige. O sistema está sendo atualmente aplicado a um *corpus* de 5.078 sonetos, decassílabos, dos séculos XVI e XVII.

O *corpus* com o qual as pesquisas estão sendo realizadas está codificado em TEI, para que ele possa ser reutilizado em outras pesquisas. O sistema de escansão em si recebe simplesmente a sequência de caracteres de um único verso, sem nenhuma marcação adicional.

Um módulo baseado em regras faz a separação silábica e a detecção da sílaba tônica. A exemplo dos trabalhos anteriores, a descrição dessas duas etapas é extremamente lacônica. A ferramenta utiliza um sistema externo de rotulagem de categoria gramatical. Com base no resultado dessa rotulagem, palavras de certas categorias, como os clíticos, são marcadas como átonas, independentemente da ortografia.

Se a separação silábica produzir exatamente 10 sílabas métricas, o sistema considera que a escansão está completa. Caso haja mais ou menos sílabas, o sistema utiliza regras para encontrar quais sinalefas, sinéreses e diéreses são aplicáveis; dialefas não o são porque a separação silábica inicial não faz sinalefas. Detalhes sobre tais regras não estão disponíveis. Os fenômenos encontrados são ordenados de mais naturais a menos naturais e então são sequencialmente aplicados — para aumentar a quantidade de sílabas, as diéreses; para diminuí-la, as sinalefas e as sinéreses.

Se a aplicação de todas as instâncias encontradas resultar num verso de 10 sílabas, considera-se que a resolução está completa. Caso contrário, o verso é considerado ambíguo e o sistema produz possíveis alternativas conforme diferentes combinações dos fenômenos. O esquema rítmico de cada alternativa é computado e comparado com os esquemas dos versos não ambíguos. O esquema mais comum é escolhido e a alternativa correspondente é selecionada.

No verso abaixo, escandindo pelo sistema, a decisão foi de fazer uma sinalefa com as sílabas *do* e *el*, em detrimento da também possível sinalefa entre *dre* e *He*:

1	2	3	4	5	6	7	8	9	10	11
Cuan-	do el	pa-	dre	He-	bre-	ro	nos	en-	se-	ña

O ritmo é computado diretamente a partir da escansão obtida, considerando-se o acento léxico como sendo o rítmico. O fato de a ferramenta empregar a rotulagem de categorias gramaticais colabora para que o esquema rítmico computado pela ferramenta tenha uma certa qualidade.



A última publicação (136) relata que o resultado produzido pelo sistema para 100 sonetos, totalizando 1.400 versos, foi comparado com aquele feito manualmente por dois anotadores humanos. Os versos incorretos foram 108, o que corresponde a uma taxa de acerto de 92,3%. A taxa de concordância entre os anotadores foi de 96,2%.

Nada é dito de forma explícita com relação à pontuação, mas a própria ausência de menção à pontuação nas discussões que existem sobre sinalefas e até mesmo sobre a pausa no meio dos versos são evidência de que a pontuação não tem papel importante no sistema. Nada se diz também sobre os estrangeirismos.

As dificuldades na resolução das sinalefas são frequentemente mencionadas na pesquisa. Trata-se de um trabalho bastante voltado ao processamento de um *corpus* específico, e por isso apenas os decassílabos são considerados. Mais importante é o fato de que o sistema toma decisões com base no conhecimento prévio do número de sílabas do verso. A aplicação da ferramenta a versos quaisquer teria de enfrentar, portanto, o desafio de não se saber quantas sílabas o verso deveria ter, o que pode ser particularmente complicado em poemas heterométricos.

Os autores desejam utilizar a ferramenta para fazer uma leitura distante dos poemas do *corpus*. Por esse motivo, um dos objetivos é que o *corpus* contenha sonetos não apenas dos poetas mais célebres e canônicos, mas sim de todos aqueles que participaram do contexto literário.

A utilização da ferramenta tem possibilitado também a correção do *corpus*. Em vários casos, versos que a ferramenta não conseguia escandir foram examinados e neles encontraram-se erros que puderam ser corrigidos ao cotejarem-se diferentes fontes.

### 3.3. Italiano

Foi encontrada apenas uma proposta para o italiano. É possível que a dificuldade na resolução de sinalefas, similar à encontrada em português e espanhol, e a impossibilidade de localização determinística da posição do acento nos vocábulos em italiano tenham contribuído para o baixo número de trabalhos.

O trabalho de Delmonte (84), apesar de mencionar o italiano repetidas vezes, concentra-se de fato no inglês e não apresenta resultado algum para o italiano.

### 3.3.1. O sistema de Robey

Robey (159) criou uma ferramenta para auxiliar na escansão dos versos da *Divina Comédia* de Dante. A ferramenta é semiautomática: ela propõe uma escansão ao usuário, que pode então corrigi-la ou deixá-la como está. Nada é mencionado com relação ao formato de entrada.

Apesar de não estar explícito no artigo, o sistema não faz uma divisão silábica das palavras ou do verso. Em vez disso, ele considera que grupos de grafemas vocálicos definem o centro de uma sílaba, sem tentar definir a qual sílaba pertencem as consoantes. Tal estratégia equivale a uma leitura do verso em que todas as sinéreses e sinalefas possíveis são realizadas. Algumas exceções são feitas pelo sistema, como no caso de um *i* entre dois grafemas vocálicos, que é considerado uma semivogal e portanto separa sílabas, como em *Tro[i]a*, que o sistema considera possuir duas sílabas e não apenas uma.

O sistema parte do princípio de que a maioria das palavras, tanto no italiano contemporâneo quanto naquele de Dante, são paroxítonas. Algumas heurísticas determinam que o acento está na última sílaba. Uma delas cuida das palavras terminadas em *r*, *n* ou *l*, que são consideradas oxítonas, como em *trattar*, *convien* e *figliuol*. Outra afirma que palavras terminadas em um grafema vocálico mais *i* ou em *e* mais um grafema vocálico recebem o acento na penúltimo grafema vocálico da palavra. As palavras que recebem um agudo ou grave em sua ortografia tem a posição do acento facilmente identificada, como em *perché* e *più*. Os monossílabos são considerados átonos.

No italiano, contudo, nem sempre é possível deduzir a posição do acento a partir da ortografia, especialmente nas proparoxítonas. O autor argumenta que seria pouco proveitoso adotar um dicionário externo que especificasse as exceções à regra, pois o acento lexical não coincide necessariamente com o acento rítmico. Encontrar o acento léxico, portanto, não implica determinar o ritmo do verso; logo, não é tão importante encontrar o acento léxico. O autor arrola vários exemplos para sustentar o argumento, como este:

ch'a la seconda morte ciascun grida;

Aqui, o acento lexical na última sílaba da palavra *ciascun* não coincide com o acento rítmico: o acento lexical é muito enfraquecido ante o acento da 10ª sílaba do verso.

A divisão silábica produzida pelo sistema não incorpora questões de acento e, como somente uma alternativa é gerada para cada verso, as sinalefas e sinéreses realizadas podem estar incorretas. É neste ponto que o

sistema solicita a intervenção do usuário: a escansão encontrada é exibida e o usuário pode fazer ajustes.

Os ajustes feitos pelo usuário são guardados pelo programa. Quando o usuário percebe que o programa, por exemplo, marcou incorretamente o acento da palavra proparoxítona *anima*, ele indica ao programa a posição correta do acento. Esse caso é então armazenado num banco de dados para ser reutilizado. O usuário pode fazer ajustes que refletem também na divisão silábica, adicionando casos de dialefas e diéreses. Como uma mesma palavra pode ter múltiplas realizações conforme o contexto, o sistema armazena todas, escolhendo a primeira para produzir sua escansão inicial e exibindo todas as alternativas ao usuário.

Como a escansão final é definida pelo usuário, o ritmo pode ser obtido por uma verificação das sílabas tônicas do verso. Nada é dito sobre a pontuação, mas a descrição da junção de grafemas vocálicos numa só sílaba não inclui nenhuma discussão sobre o papel da pontuação na regra, o que é uma evidência de que a pontuação é ignorada pelo programa. Quando uma escansão é sugerida ao usuário, o sistema exibe, entretanto, a pontuação.

Nenhuma tentativa de avaliar a acurácia da escansão apresentada ao usuário é feita, possivelmente porque o objetivo do sistema nunca foi produzir um resultado 100% automático. A principal limitação do sistema é a necessidade de cooperação de humanos. Em particular, a determinação de quais sinalefas e sinéreses precisam ser feitas para a correta escansão de um verso é uma dificuldade que o sistema transfere ao usuário. O autor comenta que Dante, diferentemente de outros poetas mais rigorosos, requer que as mesmas palavras sejam pronunciadas ora com diérese, ora com sinérese.

O sistema auxilia na coleta de informações estilísticas, como a utilização de diérese e dialefa, que podem servir para caracterizar temas ou partes do poema de Dante. Talvez devido às limitações da época em que foi realizado ou ao escopo do trabalho, que se limita à *Divina Comédia*, não houve tentativa de resolver sinalefas e sinéreses. O fato de a posição do acento não poder ser deterministicamente localizado também não é um incentivo.

### 3.4. Francês

O francês, embora uma língua românica, apresenta uma fonologia com propriedades bastante diferentes daquelas do português, do espanhol e do

italiano. A poesia francesa reflete tais diferenças: as questões de sinalefa e sinérese, por exemplo, são bastante simplificadas em francês, enquanto problemas como a classificação do grafema *e* no fim de palavras são particulares ao francês. É comum, por exemplo, que os trabalhos em francês falem de uma palavra que *possui* uma sinérese num dado autor; comentam até mesmo sobre a criação de dicionários de sinérese. Em português, por outro lado, as sinéreses e diéreses no interior de uma palavra estão sempre sujeitas a serem alteradas, numa mesma época, num mesmo poeta e num mesmo verso.

Os trabalhos de análise métrica em francês começaram ainda na década de 80 com Roubaud (162, 163); não se conseguiu, contudo, acesso ao seu trabalho. No fim da década de 90, surgiu o projeto *Metromètre* e, a partir da década de 2010, o *Anamètre*; os dois projetos são descritos a seguir.

### 3.4.1. *Métromètre*

Valérie Beaudouin desenvolveu o sistema *Métromètre* (ou seja, *metrômetro*, que mede as medidas), capaz de analisar certos versos em francês (54). A pesquisa que produziu o sistema tinha a intenção de analisar versos alexandrinos do período clássico (entre 1630 e 1830, aproximadamente), o que acabou limitando o próprio sistema a este tipo de verso e, mais do que isso, a esse período de tempo. A ideia não foi, portanto, produzir um sistema que analisasse versos em francês de forma geral, mas sim analisar estatisticamente obras de importantes escritores do período clássico do alexandrino francês.

O sistema *Métromètre* faz a transcrição fonética do verso após uma etapa de análise sintática, o que facilita a marcação da sílaba tônica das palavras. Em francês, a sílaba tônica é a última (como em *épée*) ou, pelo menos na poesia, a penúltima (*attendre*), de sorte que encontrar a sílaba tônica de uma palavra não é o desafio. O mais difícil é saber quais palavras são átonas, que não portam nenhum acento quando integradas no verso. Neste momento é que a análise sintática ampara a transcrição fonética: o *Métromètre* considera tônicos somente os verbos, substantivos, adjetivos e advérbios; as demais são átonas.

Para dividir o verso nas 12 sílabas do alexandrino, o sistema trata três situações que podem influenciar o número de vogais plenas na transcrição fonética: a diérese, o *e* mudo e a ligação.

O sistema precisa saber quando a diérese ou a sinérese é aplicável, já que a escolha entre elas corresponde à diferença de uma sílaba na divi-

são em sílabas. Considere-se o seguinte verso, escandido e exemplificado pela autora (51):

Et que j'entende dire aux peuples indiens

1 2 3 4 5 6 7 8 9 10 11 12  
e:kə:|jã:tã:də|dir|o:pø:|pləz|ẽ:|di|ẽ

A última palavra, na pronúncia corrente do francês, consiste de duas sílabas, /ẽ'djẽ/, mas, para que as 12 sílabas do alexandrino sejam obtidas, o Métromètre deve realizar a diérese, dividindo-a em três sílabas, /ẽ.di'ẽ/. Casos há em que a mesma palavra sofre ora diérese, ora sinérese, conforme o autor. Beaudouin observa, por exemplo, que Racine vê duas sílabas em *hier*, enquanto Corneille e Molière veem uma.

O *e* mudo é o grafema *e* que aparece no fim de muitas palavras francesas. Este grafema, na poesia, pode ou não ser pronunciado, conforme o contexto. Em geral, quando ele se encontra antes de outra vogal ou no fim do verso, ele é elidido. O seguinte verso, exemplificado por Beaudouin (51), mostra casos de elisão e de permanência:

Refus[e] cett[e] joi(e), et s'en trouv(e) accablé(e)

Aqui, os *ee* entre colchetes são pronunciados com o som /ə/, enquanto os que estão entre parênteses não são pronunciados em absoluto.

O terceiro caso que influencia a contagem de sílabas métricas no francês é a ligação (*liaison* em francês): são as consoantes finais das palavras, que normalmente são mudas mas que podem vir a ser pronunciadas se a palavra seguinte começa por uma vogal. Desse modo, o *s* não é pronunciado em *les filles*, mas o é em *les enfants*. O número de sílabas num verso pode ser afetado pela ligação porque, sem ela, um *e* final fica exposto e pode sofrer elisão, como no verso seguinte, mencionado por Beaudouin (51):

Me nourrissant de fiel, de larm[es] abreuvée,

Aqui, o *s* final de *larmes* precisa ser pronunciado, de outro modo o *e* que o precede seria elidido e o verso contaria 11 sílabas apenas.

O Métromètre não considera mais de uma alternativa para cada verso: ele utiliza seu conjunto de regras para produzir uma única escansão final. Ademais, apesar de o sistema ter sido desenvolvido especificamente para analisar os dodecassilábicos alexandrinos, o sistema não tenta forçar os versos à métrica; as 12 sílabas ou são encontradas naturalmente ou o verso é escandido incorretamente. Tal característica do sistema foi mesmo explorada de forma vantajosa para encontrar diferenças entre autores: como o Métromètre foi originalmente preparado para processar Racine e Corneille, Beaudouin aplicou-o a Molière e examinou os versos para

os quais o sistema encontrou 11 ou 13 sílabas. Esses seriam os versos que contêm as características que distinguem Molière dos outros dois. Ela descobriu, por exemplo, que Racine e Corneille consideram que *devriez* possui 3 sílabas, ao passo que Molière conta nela apenas 2 sílabas (53).

Não é necessário fazer a análise das rimas, visto que se sabe de antemão que as rimas são emparelhadas, isto é, cada verso de número ímpar rima com o próximo. Trata-se, evidentemente, de uma premissa do alexandrino clássico francês que não se aplica a outros casos.

De posse do Métromètre, Beaudouin definiu quatro módulos para a produção de informações estilísticas sobre autores e obras (50):

- O *corpus*, codificado num formato específico que permite, entre outros, reunir versos enunciados por mais de um personagem;
- O próprio Métromètre;
- O rimário, ou seja, uma coleção de palavras que rimam umas com as outras, construída por meio do princípio transitivo de que se um verso A rima com outro B, e B rima com C, então necessariamente A rima também com C;
- As ferramentas de estatística textual, que examinam lexicalmente os versos para lhes atribuir uma classe léxico-semântica.

As informações desses quatro módulos foram armazenadas por Beaudouin numa base de dados, de forma que consultas pudessem ser facilmente feitas. A autora comparou, por exemplo, o estilo de Corneille e de Racine (52). Para fazê-lo, ela contrastou a preferência dos autores pelos diferentes padrões acentuais em cada um dos dois hemistíquios do alexandrino. Não apenas ela constatou que existe diferença entre os autores, mas que também existe diferença segundo o gênero da obra (tragédia, comédia). O trabalho de Beaudouin foi utilizado ainda por outros para realizar pesquisas estilísticas (55) e, mais recentemente, para anotar automaticamente um *corpus* como parte de uma aplicação multimídia (70).

A acurácia das escansões produzidas pelo Métromètre não foi estimada. Algumas estatísticas indicam, contudo, que o desempenho é bom: entre todos os alexandrinos processados, 99,7% tiveram a sua sexta posição marcada como fim de palavra, o que é indício de uma escansão correta.

### 3.4.2. Anamètre

O projeto Anamètre (79, 82, 81) gerencia um *corpus* que atualmente conta com mais de 500.000 versos e ferramentas para a escansão e marcação automática de versos em francês. Diferentemente do Metromètre, o Anamètre não se restringe a um tipo específico de verso. Existem trabalhos que utilizam os resultados do projeto para derivar conclusões literárias (109).

A ferramenta parte do texto escrito e codificado em XML e não faz a transcrição fonética dos versos, preocupando-se sobretudo com os núcleos vocálicos, que são encontrados por meio de heurísticas. As vogais são classificadas em três tipos: estáveis, instáveis e semivogais. As vogais instáveis são os *ee* em fim de palavra, que são subclassificados pela ferramenta em três tipos: o masculino, pronunciado sempre; o feminino, às vezes pronunciado; e o elidido, não pronunciado. Os casos de diérese são tratados por meio de um dicionário. Os autores mencionam a importância da pontuação na poesia (79), mas aparentemente a ignoram na análise automática.

Como o sistema não sabe inicialmente qual é a métrica do poema sendo examinado, é preciso um modo de determiná-la. Se os versos de um poema contêm 8 sílabas ou menos, o metro do poema é classificado como simples e nenhuma análise subsequente é feita. Se os versos possuem mais de 8 sílabas, são considerados compostos e o sistema busca definir qual é o tamanho de suas duas metades, o que equivale a dizer que ele procura pela cesura nos versos. Um conjunto de propriedades é avaliado para cada sílaba de cada verso, formando uma grade em que padrões podem ser detectados. Estas propriedades indicam, de modo geral, o quão fraca é a sílaba naquela posição. Assim, a cesura é determinada a partir da coluna em que tais propriedades estão mais ausentes, o que mostra uma posição regular nos versos em que uma sílaba forte é encontrada.

A localização da cesura, entretanto, não leva em consideração apenas a grade de propriedades. Se o algoritmo encontrar uma posição silábica em que não há nenhuma propriedade para todos os versos analisados, então de fato a grade determina o padrão métrico para o poema, por mais incomum que ele seja. Se pelo menos uma posição na grade contiver alguma propriedade, então a força da tradição entra em jogo. Através da comparação da grade com perfis pré-definidos, o sistema classifica o poema como um todo: os perfis mais típicos são 6+6 para dodecassílabos, 4+6 e 5+5 para decassílabos e 4+5 e 3+6 para eneassílabos. A possibilidade de ritmos diferentes para versos de igual medida dentro de um poema não são atualmente consideradas (81). Com o perfil do poema definido, os

versos são individualmente analisados para que determinados fenômenos sejam detectados. É possível, por exemplo, que um dodecassílabo contido num poema de perfil 6+6 tenha realização 4+4+4.

A abordagem adotada pelo projeto Anamètre com relação ao tratamento do texto escrito é parecido com as encontradas na literatura para o processamento da poesia em inglês. Um vetor de características é extraído de cada verso e a relação entre os resultados finais do programa e o texto inicial fica bastante distante. No caso do Anamètre, os núcleos vocálicos determinam o vetor, e o resultado final, que é a determinação do perfil do poema e do ritmo de cada verso, não pode ser mapeado ao texto original para, por exemplo, dividi-lo em sílabas.

Não foram realizados experimentos que permitissem avaliar a qualidade da escansão produzida pelo sistema. Por se tratar de uma ferramenta aplicável a versos quaisquer, a questão das sinéreses, por exemplo, pode influenciar na taxa de acerto. Ao que tudo indica, a ferramenta não dispõe de meios para considerar mais de uma escansão para cada verso: uma vez que uma sinérese foi feita, não há mais volta.

### 3.5. Provençal

Foi encontrada uma proposta para o provençal antigo, língua relativamente próxima ao francês; ela está descrita abaixo.

#### 3.5.1. O sistema de Rainsford e Scrivner

Rainsford e Scrivner (155) descreveram os métodos utilizados no processo de adição de marcações métricas a um *corpus* de versos escritos em provençal. O *corpus* escolhido pelos autores é a obra cujo título original é *Lo poema de Boecis*, escrita por volta do ano 1000. Ela contém 257 versos que seguem o mesmo padrão métrico: são compostos cujos hemistíquios medem 4 e 6 sílabas. São verdadeiramente compostos, já que após a 4ª sílaba pode haver uma não contada, como no seguinte exemplo, dado pelos autores:

1    2    3    4    5    6    7    8    9    10   11  
Nos|jo-|ve|om-|ne|quan-|dius|que|nos|es-|tam

Os autores utilizaram um algoritmo para a silabação automática dos versos, mas a posição da sílaba tônica foi marcada manualmente ao mesmo tempo em que a divisão silábica era validada pelo usuário. Também foram manualmente anotadas as potenciais sinéreses e sinalefas.



O algoritmo de escansão não utiliza as anotações de sinéreses e sinalefas de forma determinística. Ele tenta todas as diferentes combinações de aplicação e não aplicação dos dois metaplasmos até que um verso de 10 sílabas poéticas em que a 4ª e a 6ª são acentuadas seja encontrado. Se tal critério não é atingido, o verso é considerado inescandível.

Não foi feita uma avaliação formal da acurácia da escansão produzida pelo sistema, mas os autores argumentam que ela não precisa ser manualmente corrigida porque todas as regras de versificação do provençal são levadas em conta. Dos 257 versos, 25 foram considerados inescandíveis; os autores constataram que tais versos são de fato exceções no texto.

### 3.6. Outras línguas

Esta seção agrupa todas as línguas que não tem proximidade imediata com o português. As línguas europeias estão descritas cada uma em sua subseção; as demais são aqui brevemente mencionadas.

Kouloughli discute perspectivas para a análise de poesia em árabe, sem detalhar ferramentas concretas (118). Alnagdawi descreveu uma maneira de encontrar a métrica de poemas em árabe através de gramáticas livres de contexto (37). Ayeche utilizou redes neurais artificiais para o reconhecimento da métrica de poemas em árabe (41). Almuhareb definiu algumas técnicas de identificação de padrões poéticos em árabe com vistas à extração de poemas (36). Kurt e Kara (120) introduziram um algoritmo para a detecção, em turco, de poemas que seguem a métrica Diwan, herdada do árabe.

Wujastyk (171) produziu um programa de computador escrito em SNOBOL 4 capaz de escandir versos em sânscrito. Na década de 80, Mayrhofer apresentou sua própria abordagem para a análise de versos numa língua derivada do sânscrito clássico (127). Ousaka, na década de 90, utilizou métodos computacionais para analisar poemas canônicos escritos numa versão do sânscrito (172, 173). Mais recentemente, Rama apresentou um método para a classificação métrica de versos em sânscrito (157). Rakshit ocupou-se da análise estilística e consequente atribuição de autoria a poemas em bangla, uma língua derivada do sânscrito (156).

#### 3.6.1. Inglês

No inglês a escansão consiste, de modo geral, em atribuir às sílabas de cada verso o qualificativo de fortes ou fracas. Fenômenos equivalentes a sina-

lefas e sinéreses existem, mas têm uma importância muito menor do que em português. Os trabalhos são numerosos e cobrem uma faixa temporal que vai de *Beowulf*, escrito antes do ano 1000, e os dias atuais, passando por Chaucer, do século XIV, e Shakespeare, que nasceu em 1564. A seguir serão brevemente descritas propostas de análise de poesia em inglês, com destaque para algumas ferramentas.

Donow (86) utilizou computadores para fazer análises estilísticas com base estatística de versos de Shakespeare. O número de sílabas de cada palavra contida no poema e também seu padrão acentual, como trocaico ou anapéstico, precisava ser manualmente armazenado num banco de dados. Foley (91) inseriu os 3.182 versos de *Beowulf* em um banco de dados de forma codificada. O trabalho do computador resumiu-se à análise estatística dos dados digitados com vistas à comprovação de certas hipóteses linguísticas.

Barber e Barber (46, 47) construíram manualmente um banco de dados com determinadas características fonéticas de 15.942 versos de Chaucer e utilizaram o computador para obter conclusões a respeito da pronúncia do *e* final de uma classe de palavras. Barquist e Shie (48) criaram um banco de dados com certas características de cada verso de *Beowulf* e produziram uma série de análises estatísticas sobre a aliteração, uma propriedade fundamental da poesia em inglês antigo. Hayward (107, 108) definiu um sistema que, com base em algumas características manualmente encontradas das sílabas de um verso, encontra graus de acentuação para cada uma; com este sistema ele analisou 1.000 versos de 10 poetas diferentes para estudá-los estilisticamente.

Plamondon (150) buscou analisar a poesia em inglês de forma fonética para então identificar o ritmo utilizado nos versos (151). Hartman (105) apresentou Scandroid, ferramenta de código aberto capaz de escanear versos em inglês; uma descrição mais científica não está disponível. Tizhoosh e Dara (169) buscaram maneiras de distinguir prosa de poesia; para fazê-lo, extraíram características como a rima e o ritmo do texto e treinaram classificadores, atingindo acurácia acima de 90%. Kavanagh (117) analisou rimas e padrões nos versos em inglês com um sistema baseado em regras.

Genzel, Uszkoreit e Och (97) descreveram uma maneira de produzir versos métricos ao traduzir poemas: as traduções são restritas a formas rítmicas. Greene, Bodrumlu e Knight (103) utilizaram aprendizado supervisionado para criar um sistema de escansão utilizável na tradução e na geração automática de poesia; o *corpus* utilizado como conjunto de treina-

mento foram os sonetos de Shakespeare. Com o objetivo de visualizar características fonéticas na poesia, McCurdy produziu o RhymeDesign (129) e o Poemage (128).

**O sistema de Hidley.** Hidley (110) construiu um sistema para a análise semiautomática dos versos de *Beowulf*. O sistema não é completamente automático pois o usuário progressivamente valida os resultados apresentados pelo sistema, que em alguma medida aprende com as correções.

O sistema faz a divisão silábica com base em determinados padrões encontrados nas palavras. O acento das palavras não é reconhecido individualmente; os versos sabidamente seguem determinadas regras rítmicas e é com base nelas que a realização acentual particular daquele verso é encontrada. O sistema faz também a análise sintática dos versos, que auxilia na divisão silábica e na localização dos acentos.

**O sistema de Hirjee.** Hirjee (111), com base em seus trabalhos anteriores na análise de letras musicais (112, 113), propôs uma ferramenta de escansão de poesia. O sistema emprega em várias etapas transcrição fonética e divisão silábica, efetuadas por módulo externo adaptado. Através de um processo iterativo, o autor criou um banco de versos com padrões rítmicos conhecidos e alinhou as sílabas das palavras dos versos aos padrões, associando assim a cada palavra a probabilidade de ela estar acentuada nos poemas. A palavra *for*, por exemplo, aparece acentuada 25% das vezes; a palavra *sound*, 98%.

Os versos de um poema a ser analisado são transcritos foneticamente e convertidos em padrões rítmicos probabilísticos, que podem ser comparados a padrões pré-definidos. A escansão final é exibida graficamente na ferramenta, que destaca as sílabas acentuadas.

**ZeusScansion.** Agirrezabal, Arrieta, Astigarraga e Hulden (34) apresentaram ZeusScansion. O sistema emprega dicionários para determinar a localização do acento tônico das palavras. Ao concatená-las para formar o padrão acentual do verso em si, o sistema também realiza a análise sintática e a consequente aplicação das seguintes regras: *a)* o acento primário das palavras de classe aberta são sílabas fortes no verso; *b)* também são fortes os acentos secundários das palavras polissilábicas e o acento primário das palavras polissilábicas de classe fechada. Se uma palavra não consta no dicionário de pronúncias, então o programa busca e utiliza a palavra mais próxima segundo uma heurística.

Para avaliar a ferramenta, um *corpus* com 759 versos foi escandido manualmente. A cada verso múltiplos padrões rítmicos aceitáveis foram associados. Em apenas 199 casos o padrão rítmico encontrado pelo programa estava entre os definidos como aceitáveis. Como exemplo, o sistema produziu a seguinte escansão incorreta para o verso de Wallace Stevens:

1 2 3 4 5 6 7 8 9 10  
I | wish | that | I | might | be | a | think- | ing | stone

A divisão silábica é a presumida, já que os autores se concentram no padrão rítmico em si. A escansão teria sido considerada correta se a palavra *I* estivesse acentuada.

**SPARSAR.** Delmonte (83) criou o SPARSAR, um sistema de análise rítmica e rímica de poemas. Sua abordagem como um todo é independente de idioma, mas sua descrição se concentra mais no inglês e seus experimentos o utilizam exclusivamente. O sistema emprega vários dicionários para obter a transcrição fonética e a divisão silábica de cada palavra individualmente; palavras não encontradas são processadas por um módulo prosódico, que presumivelmente pode ser configurado com palavras não escritas por extenso. A análise sintática é empregada para distinguir, por exemplo, entre casos em que a mesma palavra pode ser átona ou tônica conforme o contexto, como *there*.

A descrição do sistema está bastante abreviada e truncada, mas ao que tudo indica a escansão propriamente dita funciona do seguinte modo. O SPARSAR associa às sílabas dos versos um qualificativo inicial de tônica ou átona com base na transcrição fonética. As sílabas podem se enfraquecer se estiverem num grupo de tônicas. O sistema produz estatísticas acerca, entre outros, do número de sílabas de cada verso do poema, mas essa informação não é utilizada para alterar o esquema rítmico já determinado para os versos.

As análises produzidas podem ser exibidas graficamente (84). Com base numa das visualizações pode-se inferir a seguinte escansão para os dois primeiros versos do primeiro soneto de *Shakespeare*, em que as sílabas sublinhadas são fortes:

1 2 3 4 5 6 7 8 9 10 11  
From | fai- | rest | crea- | tures | we | de- | si- | re | in- | crease,  
That | there- | by | beaut- | y's | rose | might | ne- | ver | die, |

Nota-se que o primeiro verso acima está incorretamente escandido, visto tratar-se de um pentâmetro iâmbico.

### 3.6.2. Alemão

A poesia do alemão, como a do inglês, preocupa-se sobretudo com a formação de pés que alternam sílabas fortes e fracas. Ainda no início da década de 80, Chisholm propôs métodos para análises de versos em alemão (72) que utilizam transcrição fonética para a extração de características estilísticas. Mais recentemente, Estes e Hench (89) apresentaram uma abordagem supervisionada para classificar sílabas métricas em alto-alemão médio.

**Metricalizer.** O Metricalizer é um sistema desenvolvido primariamente por Klemens Bobenhausen desde pelo menos 2009. O sistema não tem se mantido estático desde sua concepção, de forma que esta seção tomou por base dois momentos distintos de sua evolução. O primeiro momento é mais antigo e é representado por dois artigos (63, 62) que descrevem as etapas iniciais do sistema. O segundo é mais recente e, além de conter uma base teórica mais fortificada, descreve etapas mais avançadas (64).

A ferramenta faz a divisão silábica das palavras, ainda que sem a intenção de encontrar a fronteira exata, já que ela não é necessária à escansão. A localização do acento é feito por meio de heurísticas; por exemplo, a primeira sílaba ou uma sílaba que contenha o grafema *ü* geralmente são acentuadas. A seguir são consideradas regras de eufonia, que dizem, por exemplo, que não devem aparecer duas sílabas tônicas adjacentes no verso. Assim, o processo de determinação das sílabas tônicas começa de forma estritamente lexical, mas, se necessário, se espalha pelas sílabas vizinhas do verso e pode até mesmo considerar versos adjacentes através de algumas regras de analogia.

Os versos são classificados pelo Metricalizer com base num conjunto de padrões de referência e numa definição de complexidade métrica, que estabelece um critério de proximidade entre um verso específico e o padrão de referência. A noção de complexidade métrica é estendida também para o poema como um todo para quantificar o quão aderente o poema é aos padrões encontrados.

Um experimento foi conduzido com 153 poemas. Após a eliminação de dois poemas por causa de problemas relacionados a apóstrofes nas edições, o Metricalizer corretamente identificou o padrão rítmico de 94% dos poemas.

### 3.6.3. Russo

Breydo realizou um trabalho pioneiro de escansão automática de versos em russo (68). Ainda na década de 90, Adams e Birnbaum (33) propuseram um sistema modular para a análise de versos em russo, que teve reflexos recentemente (59). O renomado programa Starling foi estendido para que versos em russo pudessem ser analisados (119).

A principal referência em russo tem sido Pilshchikov (146, 147, 148, 149, 65), apesar de haver esforços também de Barachnin (43, 45, 44) e de Almenova (35).

### 3.6.4. Tcheco

Sgallová (165), em um trabalho pioneiro, coletou dados rítmicos e métricos sobre 6.466 versos do poeta František Jaromír Rubeš e desenvolveu um método para codificá-los em cartões perfurados. Ela realizou ainda análises estatísticas e estilísticas. Seu método de escansão, contudo, era manual. Após um hiato, ela continuou sua pesquisa, produzindo um tesouro de versos tchecos com cerca de 5.000 poemas (164), ainda utilizando o computador somente para armazenamento e análise.

**O sistema KVĚTA.** Ibrahim e Plecháč (114) e Plecháč (152) desenvolveram o sistema KVĚTA, que tem por objetivo analisar métrica e ritmicamente poemas escritos na língua tcheca. Ela auxiliou na marcação de um *corpus* poético tcheco que conta com mais de 2,5 milhões de versos (153). O sistema recebe como entrada um poema cujas palavras devem conter anotações de lematização e de morfologia. A anotação de morfologia precisa apenas especificar a categoria gramatical da palavra e seu qualificativo de negação.

O KVĚTA aplica aos versos uma série de regras, que os vão transformando pouco a pouco numa transcrição fonética. As regras levam em consideração fenômenos de assimilação, como é o caso do grafema *k* em *kdy*, que se torna sonoro na pronúncia. Entre as dificuldades de transcrição destacam-se o tratamento das sequências *au*, *ou* e *eu*, que podem ser realizadas como ditongos ou como hiatos. Em determinados contextos ortográficos, existem regras que determinam com grande acurácia qual realização é a correta; contudo, há situações em que não é possível fazer a distinção por meio de regras. Nesse caso o sistema consulta um dicionário que especifica a realização correta para cada forma.

As palavras estrangeiras constituem um caso mais complexo. Além de dificuldades semelhantes às aquelas que se enfrentam no português, os autores apontam uma que é particular ao tcheco: palavras estrangeiras não completamente vernaculizadas podem manter sua forma escrita constante enquanto a pronúncia varia para expressar os casos da língua tcheca. Um exemplo mencionado é o nome próprio *Shakespeare*, que, escrito desta maneira, no nominativo é pronunciado /ʃejkspi:r/ e no vocativo, /ʃejkspi:re/. O sistema utiliza um dicionário para transcrever palavras estrangeiras. A detecção das palavras se faz por meio de certas regras, que procuram por grafemas ausentes em tcheco, como o *cê-cedilha*, ou por certas combinações, como *ais* e *aet*.

A partir da transcrição fonética, o sistema cria uma estrutura bidimensional que corresponde aos versos do poema e às sílabas dos versos. As sílabas são encontradas a partir dos seus núcleos na transcrição fonética, que correspondem a vogais breves, longas, ditongos e algumas consoantes silábicas, como /l/ e /ɾ/. A cada sílaba nesta estrutura é calculada uma série de atributos booleanos que ajudam a caracterizá-la: *a)* se a sílaba está no início de uma palavra; *b)* se está no fim; *c)* se é uma palavra gramatical ou lexical; *d)* se é uma preposição; *e)* se é precedida de uma preposição monossilábica; *f)* se é precedida de uma sílaba em início de palavra; *g)* se é seguida de uma sílaba contendo vogal ou ditongo.

Com base no número de sílabas dos versos do poema, o sistema gera uma série de possibilidades rítmicas. A cada possibilidade rítmica do poema são associadas adaptações aos versos individuais, já que nem todos os versos podem ter o mesmo número de sílabas. Por exemplo, para uma quadra com versos alternando entre 6 e 3 sílabas, o sistema geraria padrões rítmicos correspondentes ao tamanho máximo dos versos, 6 sílabas. Um dos padrões rítmicos corresponderia a uma série de troqueus (ou seja, 101010), que seria adaptado aos versos de 3 sílabas removendo-se do fim a quantidade supérflua (resultando em 101).

O sistema KVĚTA gera atualmente padrões rítmicos de quatro categorias: *a)* os clássicos, derivados de iampos, troqueus, dátilos e anfíbracos; *b)* imitações de metros silábicos quantitativos, como é o caso da estrofe sáfica; *c)* outras imitações, incluídos o pentâmetro e o hexâmetro datílico; e *d)* os ritmos que contém um *radif*, ou seja, uma palavra que é adicionada ao fim de determinados versos, aumentando-lhes o tamanho.

O sistema passa então a comparar os padrões encontrados nos versos em si com aqueles gerados. Inicialmente, uma ideia de índice métrico, adaptada do trabalho de Pilshchikov e Starostin (147) era utilizada. Tal in-

dice é calculado através de uma série de regras que partem de um valor inicial e o vão diminuindo conforme inadequações são encontradas. Os coeficientes utilizados no cálculo do índice métrico foram determinados de forma subjetiva. Mais recentemente, esse índice está sendo substituído por um coeficiente métrico que utiliza parâmetros probabilísticos derivados do *corpus* revisado de que os pesquisadores dispõem atualmente.

O melhor índice ou coeficiente encontrado para o verso entre os padrões rítmicos gerados determina a sua classificação, contanto que certas condições sejam atendidas. Entre outras, o melhor índice ou coeficiente precisa estar acima de um certo valor mínimo. Caso a classificação automática não seja possível, o sistema repassa a responsabilidade de classificação ao usuário.

A acurácia do sistema foi medida com relação ao *corpus* de referência do grupo de pesquisa. Esse *corpus* foi anotado de forma semiautomática: a própria ferramenta KVĚTA foi utilizada na sua anotação métrica, mas, como visto há pouco, casos mais difíceis são resolvidos pelo usuário. Para estimar o quão confiável é o *corpus*, os pesquisadores avaliaram manualmente 300 poemas, totalizando 25.779 versos, e constataram que 99,97% tinham a anotação métrica correta.

O sistema processou os 2.336.435 versos do *corpus*, produzindo resultados incorretos em 108.979 dos casos, o que resulta numa taxa de acerto de 95,34%. Com a utilização do índice de métrica antigo, esta taxa cai para 94,88%.

### 3.6.5. Latim e grego

Estas duas línguas clássicas possuem tradições poéticas similares, tanto por causa da fonologia similar quanto da influência cultural do grego sobre o latim. Propostas que buscam tratar os dois idiomas simultaneamente existem em número suficiente para que sejam aqui apresentadas na mesma subseção. Começando cedo, ainda na década de 60, Jones (116) propôs mecanismos para a codificação de hexâmetros em computador, e Greenberg (102) apresentou sua proposta para escansão completamente automática do hexâmetro datílico, atingindo 92% de precisão. Bem mais recentemente, Fusi vem trabalhando com o verso clássico (95, 94), produzindo um sistema especialista (93) e um arcabouço (92).

Especificamente para o latim, os estudos começaram desde a década de 60 com Ott (142, 143, 144). Na década de 70 o interesse continuou com a pesquisa de Berschin (57) e de Jiménez (115). Também a década de 80



viu trabalhos na área (133). Mais recentemente, Deufert tem trabalhado com a escansão do latim (85, 61).

Especificamente para o grego houve menos propostas. Na década de 70, Évrard (90) propôs métodos para a escansão automática do hexâmetro grego. Mais recentemente, Papakitsos (145) também trabalhou com o hexâmetro grego.

### 3.7. Considerações finais

Os trabalhos apresentados neste capítulo mostram que o interesse na análise métrica automática existe já há bastante tempo, apesar de que os esforços de pesquisa com frequência dependeram da obstinação de um único indivíduo ou laboratório. Nos anos recentes, parece haver um grande aumento no interesse pela área, provavelmente um resultado da popularização das humanidades digitais.

Os desafios das ferramentas de análise de poesia variam conforme a língua. As línguas mais próximas ao português, como o espanhol e o italiano, precisam enfrentar a presença de fenômenos como sinalefas e sínéreses, que devem ser avaliados caso a caso e muitas vezes num contexto que vai além do verso. Ferramentas que analisam mais de uma língua são bastante excepcionais e dependem justamente de uma certa proximidade, como é o caso do latim e do grego. Línguas como o português e o espanhol, por exemplo, são candidatas a serem tratadas por uma única ferramenta.

A Tabela 3.3 lista as ferramentas encontradas na literatura para a escansão de poemas no português e em línguas próximas. As colunas *Estrangeirismos* e *Expansões* se referem à capacidade do sistema de tratar automaticamente textos com palavras em língua estrangeira e com palavras não escritas por extenso. Vários sistemas, inclusive o Aoidos, possibilitam que tais casos sejam tratados, mas não de forma automática. A coluna *Pontuação* indica se a ferramenta utiliza a pontuação presente nos versos de alguma forma; como se vê, a regra entre os sistemas é descartá-la. A coluna *Fonética* mostra se o sistema trabalha com a transcrição fonética dos versos. As duas abordagens em português o fazem; uma das duas em francês também; a única em espanhol que emprega transcrição fonética o faz para a representação visual de sons. Em *Sintaxe*, indica-se se a abordagem adota algum tipo de análise sintática com o propósito de escandir versos.

A coluna *Contexto* faz menção à capacidade do sistema de considerar mais de uma alternativa para a escansão do verso, conforme seu contexto. Em francês, esta habilidade não é tão importante; mas em português,

Tabela 3.3.: Tabela comparativa das ferramentas propostas para o português e línguas próximas. Os símbolos possuem o seguinte significado: ✖ = ausente; ✔ = presente. A coluna *Ref.* contém o número da referência bibliográfica.

Língua	Ano	Ref.	Ferramenta	Estrangeirismos	Expansões	Pontuação	Fonética	Sintaxe	Contexto	Geral	Acurácia	N1	N2
Francês	1996	(51)	Metromètre	✖	✖	✔	✔	✔	✖	✖	-	-	79.456
Espanhol	1998	(160)	Colors of Poetry	✖	✖	✔	✔	✖	✖	✔	-	-	-
Espanhol	2000	(98)	-	✖	✖	✖	✖	✖	✖	✖	88,7	896	896
Português	2008	(124)	SAEP	✖	✖	✖	✔	✖	✖	✔	≤82,2	197	197
Francês	2011	(81)	Anamètre	✖	✖	✖	✖	✖	✖	✔	-	-	540.000
Provençal	2014	(155)	-	✖	✖	✖	✖	✖	✔	✖	-	-	257
Espanhol	2015	(135)	-	✖	✖	✖	✖	✔	✔	✖	96,2	1.400	71.092
Português	2016	(131)	Aoidos	✖	✖	✖	✔	✖	✔	✔	99,0	1.080	107.438

espanhol e provençal ela é essencial para que muitos versos sejam escandidos corretamente. Existem poucas taxas de acerto que se podem comparar, mas a taxa de acerto da ferramenta de Gervás está mais de 7 pontos percentuais abaixo da de Colorado; o algoritmo de resolução de sinalefas e sínéreses deste último certamente é uma das razões. Igualmente o Aoidos consegue uma taxa muito superior àquela do SAEP, o que parcialmente se deve às características determinísticas deste último.

A coluna *Geral* refere-se à aplicabilidade do sistema: se ele pode escandir versos quaisquer ou se ele é restrito a um tipo específico de verso. Um sistema geral precisa analisar os versos individualmente e também o poema como um todo, o que aumenta a sua complexidade. A coluna *Acurácia* mostra a taxa de acerto no principal experimento de validação encontrado; o número de versos utilizados no experimento é dado pela coluna *N1*. Por fim, a coluna *N2* mostra o número total de versos aos quais o sistema foi aplicado.

A acurácia do sistema SAEP é dada como ≤82,2 na tabela porque os autores não verificaram a escansão dos versos, mas somente o número de sílabas poéticas. A taxa de acerto de 82,2%, portanto, representa um teto, atingível caso as escansões estejam corretas sempre que o sistema acertou o número de sílabas.

## 4. CORPORA

1	2	3	4	5	6	7	8
Eu	não	me	que-	ro e-	men-	dar,	
pois	fa-	ço	ver-	so-	s em	ri-	mas,
e à-	s u-	nba-	da-	s os	su-	jei-	to,
de	quem	os	cor-	ta, e	be-	lis-	ca.

Gregório de Matos

Este capítulo apresenta todos os *corpora* utilizados no desenvolvimento e nos experimentos descritos nesta tese. Tais *corpora* estão agrupados em três classes, detalhadas nas seções seguintes: os *corpora* lexicais, cujas palavras, consideradas individualmente, são material relevante às etapas de localização do acento e silabação de palavras do Aoidos; os *corpora* poéticos principais, que servem de base ao estabelecimento das regras e à validação do sistema como um todo; e os *corpora* poéticos secundários, utilizados especificamente num dos casos de uso apresentados no Capítulo 6.

A maior parte dos *corpora* foi derivada a partir de edições disponíveis na Literatura Digital<sup>1</sup>, a biblioteca de literaturas de língua portuguesa mantida pelos laboratórios NUPILL (Núcleo de Pesquisas em Informática, Literatura e Linguística) e LaPeSD (Laboratório de Pesquisa em Sistemas Distribuídos) da UFSC (Universidade Federal de Santa Catarina).

### 4.1. *Corpora* lexicais

O *corpus* HOU é composto de todas as 117.973 entradas do dicionário Houaiss em sua versão 2009.6, de julho de 2013 (*sic*). Cada entrada define, em geral, uma palavra, símbolo ou abreviação. O objetivo deste *corpus* é cobrir grande parte das possibilidades permitidas pela ortografia do português, sem considerar a frequência de uso das palavras, já que palavras raras ou comuns têm igual presença. Foi um *corpus* útil, portanto, durante o desenvolvimento do sistema, já que fornece exemplos de palavras com combinações pouco comuns de grafemas, como *juuna* (um tipo de arbusto) e

<sup>1</sup><http://www.literaturabrasileira.ufsc.br/>

*baniua* (membro de um certo grupo indígena), que precisam ser tratadas pelo sistema em suas regras. Deste *corpus* foi derivado outro, denominado HOU10, com 10.000 palavras selecionadas aleatoriamente a partir do original. O *corpus* menor foi utilizado nas avaliações manuais do Capítulo 6.

O segundo *corpus* lexical, LIT, foi obtido a partir da Literatura Digital, que conta atualmente com 4.540 arquivos digitalizados, dos quais mais de 1.100 em formato HTML. Para a composição do *corpus* LIT foram selecionadas, de forma aleatória, 66 obras escritas em português de acordo com as normas da ortografia contemporânea, em verso e em prosa, de 38 autores diferentes, totalizando quase 2,8 milhões de palavras, das quais pouco mais de 240 mil únicas. Este é um *corpus* em que a frequência com que as palavras são utilizadas na prática torna-se, evidentemente, diferente daquela do HOU. Foi derivado a partir do LIT um *corpus* menor, LIT30, desta vez com 30.000 palavras aleatoriamente selecionadas, totalizando 8.023 únicas, também com o propósito de fornecer material para os experimentos do Capítulo 6.

## 4.2. *Corpora* poéticos principais

Os *corpora* poéticos principais descritos nesta seção têm o objetivo de fornecer material para o desenvolvimento das regras do sistema de escansão e também para a sua validação. Quanto maior for a quantidade e variedade de versos submetidos ao sistema, maior será a abrangência das regras e mais confiável a validação. Por outro lado, o custo envolvido no estabelecimento de *corpora* poéticos é proporcional não apenas à quantidade de versos, mas também à variedade das fontes: é mais fácil criar *scripts* para tratar um único arquivo HTML grande do que vários pequenos. Tendo em vista tais circunstâncias, os seguintes critérios foram definidos para a seleção dos *corpora* poéticos principais:

- A obra deve estar disponível na Literatura Digital em formato HTML. Deste modo tem-se acesso a um grande conjunto de obras que seguem um padrão homogêneo de codificação.
- Ela deve estar escrita na língua portuguesa. Versos ou poemas escritos em língua estrangeira, bastante comuns em obras poéticas, podem estar presentes, mas devem constituir a minoria absoluta.
- Ela deve adotar a ortografia contemporânea do português, entendida como aquela praticada no Brasil desde a década de 1970. Obras

mais antigas podem ser usadas, desde que sua ortografia esteja atualizada.

- Ela deve conter ao menos 4.000 versos, para que seja encontrando um equilíbrio entre poucas obras com muitos versos e muitas obras com poucos versos. Este número é bastante arbitrário, mas na prática averiguou-se que há uma boa quantidade de obras que passam deste limiar.
- É aceita no máximo uma obra por autor, para manter uma certa variedade.
- Os arquivos que compõem a obra na Literatura Digital devem corresponder exatamente um livro físico, e devem conter todos os poemas nele presentes.
- Deve ser possível acessar pelo menos uma edição impressa da obra, seja através de livros escaneados ou em papel.
- As obras, em seu conjunto, devem cobrir uma grande faixa temporal, para garantir que o sistema não fique restrito às convenções de apenas um período.
- A obra não pode conter qualquer quantidade de versos livres. Poemas em que o tamanho dos versos varia livremente não eliminam a obra, contanto que cada verso seja individualmente métrico.

Com base nos critérios acima, foram selecionadas obras suficientes para que fossem atingidos no mínimo 100.000 versos. As obras selecionadas constam da Tabela 4.1, em que estão ordenadas pela data de morte do autor. As obras cobrem mais de três séculos de poesia, do prestigioso épico de Camões aos poemas da catarinense Delminda Silveira. Entre os *corpora* encontram-se a obra poética completa de três autores (MAT, ANJ e SIL), obras que contêm um único poema longo (CAM, DUR, GON e PIN) e livros variados (COS, MAG e FIG). A obra de Xavier Pinheiro (PIN) é excepcional, pois trata-se da única tradução entre os *corpora*.

As obras escolhidas para constituírem os *corpora* poéticos principais foram convertidas, através de procedimentos semiautomáticos, de HTML para o formato TEI (*Text Encoding Initiative*, baseado em XML). Esta conversão preservou somente títulos e versos, a maioria do restante do material sendo eliminado. Duas revisões sucessivas foram feitas nos *corpora*, que são aqui denominadas a revisão preliminar e a pontual.

Tabela 4.1.: *Corpora* poéticos principais, ordenados pela data de morte do autor da obra. A *Divina Comédia* de Xavier Pinheiro é uma tradução, o original sendo de Dante Alighieri.

Código	Autor	Morte	Obra
CAM	Luís de Camões	1580	Os Lusíadas
MAT	Gregório de Matos	1696	Obra Poética
DUR	Santa Rita Durão	1784	Caramuru
COS	Cláudio Manuel da Costa	1789	Obras
GON	Tomás Antônio Gonzaga	1810	Cartas Chilenas
MAG	Gonçalves de Magalhães	1882	Suspiros Poéticos e Saudades
PIN	Xavier Pinheiro	1882	Divina Comédia
ANJ	Augusto dos Anjos	1914	Obra Completa
FIG	Araújo Figueredo	1927	Poesias
SIL	Delminda Silveira	1932	Obra Completa

A *revisão preliminar* buscou encontrar faltas e excessos de versos nos *corpora*. Existem contagens de versos e tamanhos de estrofes conhecidos para várias obras, como é o caso para o *corpus* CAM: sabe-se que existem precisamente 8.816 versos arranjados em oitavas; qualquer desvio indica algum tipo de erro. Outros *corpora*, como ANJ ou SIL, não possuem características tão bem definidas e são mais passíveis de conterem versos a mais ou a menos.

A *revisão pontual* foi feita quando os *corpora* foram submetidos à análise automática. Existem problemas nas edições que podem ser detectados imediatamente: o sistema se recusa a escandir versos contendo certas más-formações, como *og* em vez de *oh* em MAG.4.6.11, ou certas palavras com grafia estrangeira, como *Macbeths* em ANJ.1.22.3. Com a escansão dos poemas terminada, heurísticas foram utilizadas para detectar possíveis problemas. Assim, poemas que o sistema considera não serem isométricos foram examinados para verificar se de fato não o são. Também foram conferidas propriedades conhecidas dos *corpora*: se existem apenas decassílabos no DUR, a escansão produzida pelo sistema precisa re-fletir esse fato.

Para cada *corpus* principal foi utilizada, além da edição da Literatura Digital, também uma de referência, necessariamente impressa (em papel ou escaneada). A edição de referência foi empregada para resolver, sempre que possível, questões encontradas durante as revisões preliminar

Tabela 4.2.: Quantidade de versos em cada *corpus* principal segundo o idioma do poema em que se encontram. O percentual se refere à proporção dos versos em português de cada *corpus* com relação ao número total de versos em português.

	Português		Espanhol	Italiano
<i>Corpus</i>	Total	%	Total	Total
CAM	8.816	8,2	555	488
MAT	33.840	31,5		
DUR	6.672	6,2		
COS	7.682	7,2		
GON	4.184	3,9		
MAG	5.637	5,2		
PIN	14.233	13,2		
ANJ	6.594	6,1		
FIG	8.233	7,7		
SIL	11.547	10,7		
Total	107.438	100,0	555	488

e pontual. Não houve, entretanto, uma conferência verso a verso entre o *corpus* derivado da edição digital e a edição de referência. Somente os versos que causaram algum problema nas revisões preliminar e pontual foram confrontados com a edição de referência. Em particular, os *corpora* podem conter uma quantidade desconhecida de erros que não influenciam nem na quantidade, nem na métrica dos versos.

A Tabela 4.2 mostra a quantidade de versos em cada um dos *corpora* principais. Poemas escritos em língua estrangeira não foram eliminados e são contados separadamente na tabela. Entretanto, o idioma de cada verso não foi classificado separadamente; o idioma dominante de cada poema foi identificado de forma semiautomática e todos os versos do poema foram contados em favor do idioma dominante. Os versos dos poemas em língua estrangeira não estão incluídos em quaisquer análises ou estatísticas desta tese, exceto quando explicitamente indicado.

Os *corpora* incluem apenas os versos que constituem o corpo propriamente dito de cada obra. Em particular, epígrafes de qualquer tipo e motes estão excluídos das contagens. Poemas ou fragmentos que aparecem em materiais introdutórios também não entram nos *corpora*, independentemente da autoria dos versos. A disposição original dos versos foi

preservada nos *corpora*, mas versos que foram quebrados apenas graficamente ficam marcados no *corpus* como constituindo uma unidade e não são contados separadamente; tais versos são chamados aqui de *emendados*. O seguinte verso decassílabo, por exemplo, foi quebrado em dois trechos:

SIL:1.116.3.5 quem é?

SIL:1.116.3.6 É tua Mãe, — a mulher Santa!

Os *corpora* não foram revisados especificamente em busca de versos emendados, mas a revisão pontual revelou de qualquer modo certa quantidade deles.

Uma caracterização mais detalhada dos *corpora* pode ser vista na Tabela 4.3, que mostra a quantidade de versos, estrofes e poemas dos *corpora*. Destacam-se os tamanhos mínimos de estrofe de dois *corpora*, COS e MAG. Este tamanho mínimo é zero porque há estrofes nestes *corpora* que contêm apenas um verso que, na verdade, é a continuação de um outro precedente. Como tais versos emendados são contados apenas uma vez, algumas estrofes podem ficar vazias; por exemplo, se uma estrofe contém apenas a última parte de um verso, ela fica vazia quando essa parte é integrada com o restante do verso. A questão permanece independentemente de o verso integral ser contado na estrofe em que começa ou termina, pois existem estrofes de um único verso que contêm tanto a primeira quanto a última parte de um verso emendado.

O número de poemas em cada *corpus* é, em alguns casos, discutível. A obra de Santa Rita Durão, *Caramuru*, é um único poema épico, constituído de 10 cantos. Na preparação dos *corpora*, as partes em que um grande poema está dividido foram consideradas elas mesmas poemas. A distinção é importante porque a unidade básica de análise do Aoidos é o poema: as características dos versos de outros poemas de uma mesma obra não influenciam diretamente na escansão. Por exemplo, as *Quadras Comerciais*, ANJ.284.1.1, de Augusto dos Anjos, seriam mais difíceis de serem analisadas pelo sistema caso estivessem dispostas em poemas separados, já que os padrões de versificação que são visíveis nos 56 versos de todas as quadras podem não ser mais identificáveis quando apenas 4 versos são considerados.

A Tabela 4.4 mostra alguns casos particulares de versos nos *corpora*. A coluna *Repetidos* conta a quantidade de versos repetidos, ou seja, se um verso aparece três vezes de forma igual no mesmo *corpus*, duas repetições são contadas. O objetivo dessa coluna é estabelecer que a quantidade total de versos de cada *corpus* não fica severamente reduzida quando versos repetidos são levados em conta. Desta forma, a tabela informa que exis-



Tabela 4.3.: Contagem de versos, estrofes e poemas em cada um dos *corpora* principais.

<i>Corpus</i>	Versos		Estrofes			Poemas			
	Total	Total	Número de versos			Total	Número de versos		
			Mín.	Máx.	Méd.		Mín.	Máx.	Méd.
CAM	8.816	1.102	8	8	8,0	10	696	1.248	881,6
MAT	33.840	3.518	1	236	9,6	694	4	959	48,8
DUR	6.672	834	8	8	8,0	10	592	736	667,2
COS	7.682	1.601	0	28	4,8	131	14	664	58,6
GON	4.184	154	4	128	27,2	14	37	434	298,9
MAG	5.637	750	0	69	7,5	56	20	256	100,7
PIN	14.233	4.811	1	3	3,0	100	115	160	142,3
ANJ	6.594	1.615	2	60	4,1	284	4	438	23,2
FIG	8.233	2.401	1	118	3,4	411	14	175	20,0
SIL	11.547	2.379	1	44	4,9	522	4	564	22,1
	107.438	19.165	0	236	5,6	2.232	4	1.248	48,1

Tabela 4.4.: Quantidade de versos repetidos, fragmentários e emendados nos *corpora* principais.

<i>Corpus</i>	Total	Repetidos		Fragmentários		Emendados	
		Total	%	Total	%	Total	%
CAM	8.816	3	0,3				
MAT	33.840	473	14,0			47	1,4
DUR	6.672						
COS	7.682	92	12,0			38	4,9
GON	4.184	30	7,2	12	2,9		
MAG	5.637	73	13,0			1	0,2
PIN	14.233	3	0,2				
ANJ	6.594	20	3,0	16	2,4	2	0,3
FIG	8.233	131	15,9				
SIL	11.547	1.286	111,4	4	0,3	6	0,5
Total	107.438	2.111	19,6	32	0,3	94	0,9

tem  $107.438 - 2.111 = 105.327$  versos sem repetição nos *corpora*. Dois versos foram considerados iguais se possuem as mesmas palavras, desprezados quaisquer sinais de pontuação.

Tabela 4.5.: Quantidade de expansões, estrangeirismos e regularizações anotadas nos *corpora* principais.

<i>Corpus</i>	Expansões		Estrangeirismos		Regularizações	
	Total	‰	Total	‰	Total	‰
CAM	1	0,1	3	0,3		
MAT	35	1,0	63	1,9	13	0,4
DUR	18	2,7	33	4,9		
COS					1	0,1
GON			4	1,0		
MAG			26	4,6	1	0,2
PIN	17	1,2	142	10,0	10	0,7
ANJ	26	3,9	84	12,7		
FIG	1	0,1				
SIL	7	0,6	9	0,8	50	4,3
Total	105	1,0	364	3,4	75	0,7

A coluna *Fragmentários* da Tabela 4.4 conta os casos em que as edições deixam claro que um determinado verso não está completo. Entram nessa contagem tanto os versos em que falta algum trecho, como ANJ.67.20.6, ou versos completamente vazios, como ANJ.67.22.1. O primeiro caso corresponde tipicamente a versos em que algum trecho não estava legível no material que originou a edição; o segundo, a versos cuja existência pode ser estabelecida no material de origem mas cujo conteúdo é inteiramente ilegível. Casos em que as edições indicam que uma quantidade não determinada de versos estão faltando não foram contabilizados. A coluna *Emendados* da mesma tabela conta os versos que estão graficamente divididos porém metricamente emendados.

A revisão pontual ressaltou versos com algumas palavras que impediam sua correta escansão. O primeiro caso se refere a palavras que não estão escritas por extenso, sejam nomes de letra, como *M* em PIN.86.43.3, ou abreviaturas, como *S.* por *São* em SIL:5.16.1.10; tais casos foram anotados nos *corpora* com a expansão textual necessária. O segundo caso são as palavras estrangeiras: o sistema é capaz de processar somente palavras que seguem a ortografia do português, o que não acontece com palavras grafadas em outro idioma, como *clown* em ANJ.170.1.1. As palavras e trechos estrangeiros foram anotados com um aportuguesamento que busca especificar uma pronúncia através da ortografia regular do português. O

terceiro e último caso corresponde a palavras cuja ortografia é questionável ou apresenta alguma particularidade, sem entretanto constituir um erro evidente; é o que acontece no verso MAT:15.12.2.2, em que *bra-*, a primeira sílaba de *braços*, está escrita sem o agudo, o que a deixa sem a tonicidade esperada. Essas formas questionáveis foram anotadas com uma regularização que elimina as ambiguidades que impediam a escansão do verso. As contagens de expansões, estrangeirismos e regularizações anotadas nos *corpora* estão apresentadas na Tabela 4.5. A necessidade destas marcações nos *corpora* para que o sistema possa escandir os versos constitui uma das limitações desta tese.

A Tabela 4.6 mostra a quantidade de versos corrigidos, faltantes, excedentes e inescandíveis encontrados nos *corpora*. Os versos faltantes e excedentes — ou seja, aqueles que a edição digital possui a mais ou a menos com relação à edição de referência — foram encontrados durante a revisão preliminar; as correções foram feitas na revisão pontual, tipicamente através do confronto entre a edição digital e a de referência, apesar de haver uma minoria de casos em que a própria edição de referência continha um erro evidente, situação em que a correção foi feita em detrimento da obra impressa. Versos que não contenham nenhum erro e que mesmo com a consultoria de um especialista não puderam ser escandidos de acordo com a métrica do poema em que se encontram — um eneassílabo quando um decassílabo era esperado, por exemplo — foram classificados como inescandíveis; a integralidade desses versos consta do Apêndice A. Todas as melhorias feitas serão repassadas à Literatura Digital.

A seguir cada *corpus* é apresentado e suas particularidades são discutidas.

**Corpus CAM.** Este *corpus* corresponde à obra *Os Lusíadas*, de Camões (8). A edição de referência utilizada foi a do Instituto Camões, de Portugal (7), diferente daquela na qual se baseou a edição digital.

A obra é extremamente regular do ponto de vista estrófico: todas as estrofes possuem exatamente 8 versos. Existem apenas 3 versos duplicados, aparentemente de forma acidental (CAM.1.44.8 e CAM.2.70.4, CAM.1.61.1 e CAM.2.77.1, CAM.3.35.8 e CAM.9.7.8).

Três estrangeirismos foram marcados no *corpus* durante a revisão pontual. O nome *David* aparece duas vezes, em CAM.1.71.4 e CAM.3.140.5, e está assim também grafado na edição de referência, possivelmente porque a intenção do poeta era a de que o *d* final fosse pronunciado. O terceiro estrangeirismo é um verso inteiro escrito em italiano:

Tabela 4.6.: Quantidade de versos corrigidos, faltantes, excedentes e inescandíveis nos *corpora* principais.

<i>Corpus</i>	Corrigidos		Faltantes		Excedentes		Inescandíveis	
	Total	%	Total	%	Total	%	Total	%
CAM	117	13,3	3	0,3	3	0,3		
MAT	240	7,1	4	0,1			45	1,3
DUR	90	13,5						
COS	89	11,6	217	28,2	253	32,9		
GON	11	2,6						
MAG	10	1,8						
PIN	101	7,1					8	0,6
ANJ	21	3,2					1	0,2
FIG	39	4,7					33	4,0
SIL	63	5,5			20	1,7	43	3,7
Total	781	7,3	224	2,1	276	2,6	130	1,2

CAM.9.78.8 Tra la spica e la man, qual muro è messo.

A única marcação de expansão foi feita no verso CAM.1.12.3, em que um *D. Fuas* deve ser lido como *Dom Fuas*.

A edição digital contém um bom número de erros (117 ou 13,3%): palavras que faltam, como *adaga* no verso CAM.2.95.6, algumas completamente alteradas, como *tamboril* em vez de *também* no verso CAM.9.34.5, outras com mais de um diacrítico, como *invisibil* em vez de *invisível* em CAM.1.65.2. Há também modernizações que facilitam o entendimento do conteúdo mas que inviabilizam metricamente o verso; em situações assim a modernização foi desfeita. É o caso de *inimigo* em vez do correto *imigo*, o que sucede 7 vezes, como no seguinte verso:

CAM.7.84.4 Inimigo da divina e humana Lei.

Também o antigo *quamanha* aparece como *que tamanha* na edição digital, o que prejudica a escansão decassílabo:

CAM.5.69.5 Ora vê, Rei, que tamanha terra andamos,

A edição digital possui ainda 3 versos espúrios, nas estrofes CAM.1.53, CAM.3.130 e CAM.8.61; todos eles são advindos de glosas que introduziam o assunto tratado e que foram, em alguma etapa do processo de digitalização, convertidas em versos. A edição digital também não continha 3 versos, CAM.1.22.8, CAM.2.19.8 e CAM.6.44.8, que foram devidamente incorporados.

**Corpus MAT.** Esta é a obra poética completa de Gregório de Matos (22). A edição de referência adotada foi a de James Amado (21), na qual também a edição digital foi baseada. Várias questões foram resolvidas consultando-se também o livro de Rogério Chociay sobre a obra poética de Gregório de Matos (74). É o único *corpus* que inclui poemas em espanhol, totalizando 555 versos naquele idioma.

As linhas que contém dois versos separados por pontilhado na edição impressa (e na digital), como no poema MAT:1.17, foram quebradas em seus dois componentes, exceto no caso do poema MAT:2.15, em que as rimas internas dos decassílabos foram separadas em duas metades, mas no *corpus* o verso ficou íntegro. O soneto em que Gregório de Matos distribuiu graficamente sílabas compartilhadas, MAT:7.5, foi disposto no *corpus* de maneira linear.

Este *corpus* possui sequências ininterruptas de versos extremamente longas, chegando a um máximo de 236 versos no poema MAT:9.5. O que sucede é que, entre outros, as quadras assonantadas de Gregório não estão graficamente separadas na edição de James Amado, todos os versos estando agrupados numa única estrofe visível. O número médio de versos por estrofe, portanto, também seria mais baixo do que aquele informado na Tabela 4.3 caso houvesse uma separação gráfica em casos como este.

Os versos duplicados deste *corpus* (473 ou 14,0%) são resultado do grande número de estribilhos empregados pelo autor; de fato, 369 ou 78% das repetições acontecem dentro de um mesmo poema. Apenas no poema MAT:9.10, há dois estribilhos que aparecem 53 vezes cada. Outras 96 repetições aparecem em dois poemas distintos e 8 em três. O fato de não haver versos fragmentários ou vazios significa apenas que eles não foram indicados nas edições impressa ou digital. O poema MAT:5.42, por exemplo, é composto de 13 estrofes de 10 versos cada, todas décimas espinelas com esquema rimático ABBAACDDC — e uma estrofe, intercalada entre as demais, de 9 versos e com esquema rimático ABBAACDDC:

MAT:5.42.6.1 Como Galeno na praça

MAT:5.42.6.2 apareceste ao luar

MAT:5.42.6.3 pobre, roubado do mar,

MAT:5.42.6.4 que era ver-te um mar de graça:

MAT:5.42.6.5 quando um pasma, e outro embaça;

MAT:5.42.6.6 não me tenham por visão,

MAT:5.42.6.7 frade sou inda em cueiros,

MAT:5.42.6.8 tornei-me aos anos primeiros,

MAT:5.42.6.9 e Bivar foi meu Jordão.

A ausência de um verso que rime em *-ão* não é indicada de nenhuma forma e portanto o verso fantasma não é contado como fragmentário ou vazio. Os 37 versos emendados (ou 1,4%) estão presentes em dois poemas, MAT:4.35 e MAT:29.23. Estes versos, todos decassílabos, estão por alguma razão divididos em duas partes na edição de James Amado, mas constituem uma unidade métrica, como atesta Chociay (74, p. 54). Em vez de juntar os dois pedaços fisicamente, a opção escolhida foi encadeá-los.

As expansões marcadas no *corpus* são formas como *D.*, que é *Dom* em MAT:39.9.4.2, e *Fr.*, que é *Frei* em MAT:22.11.1.4, além de nomes de letras no poema MAT:1.4, como na estrofe seguinte:

MAT:1.4.1.1 De dous ff se compõe

MAT:1.4.1.2 esta cidade a meu ver

MAT:1.4.1.3 um furtar, outro foder.

Os estrangeirismos anotados no *corpus* se devem sobretudo aos versos e palavras avulsas em espanhol e latim, como o seguinte:

MAT:29.98.1.18 verbi gratia, sota, e basto

As regularizações, em número de 13 ou 0,4%, são de três tipos. Há uma ocorrência da palavra *ta*, sem diacrítico e sem o sentido de contração de *te* e *a*, que precisa ser tônica, no verso seguinte:

MAT:15.5.1.1 Tomas a Lira, Orfeu divino, ta,

Este *ta* foi regularizado para *tá*. Há 7 casos em que a forma contrata *pes*, de *pese*, aparece no texto em posição final de verso. Como está escrita, sem diacrítico, esta palavra se assemelha a *lhes*, e não a *mês* ou *pés*, motivo pelo qual foi regularizada para *pés*. Por fim, em alguns poemas em que Gregório de Matos omite a última sílaba ou som de cada palavra, em 5 versos a edição impressa não utilizou diacríticos para marcar a sílaba tônica dos vocábulos parciais, como no seguinte caso:

MAT:15.12.1.2 Que se é via do Céu a via sa-

Entre as 240 correções feitas (7,1%) estão 130 casos em que *lhe* estava escrito como *Ihe* (com um *I* maiúsculo no início), claramente um resquício da etapa de OCR na digitalização dos textos. Há ainda casos de palavras que estão faltando, como *Amigo* no verso MAT:29.72.4.1, e casos em que o diacrítico não está presente, como *roido* em vez de *roído* no verso MAT:29.81.3.1. Um total de 4 versos foram inseridos no *corpus* por estarem presente na edição impressa porém não na digital, MAT:1.8.1.47, MAT:5.24.1.77, MAT:5.41.6.5 e MAT:9.10.45.7.

Os 45 versos classificados como inescandíveis são, em sua maioria, erros em códices que sobreviveram na edição de James Amado e, consequentemente, na digital. Vários casos podem ser elucidados quando ver-

sões alternativas dos poemas são encontradas *on line*, como é o seguinte verso, que deveria ser heptassílabo:

MAT:1.5.1.16 luz, sol, mortos, e vivos:

É possível encontrar fontes que afirmam que a palavra correta é *lua*, e não *luz*<sup>2</sup>. Casos há também em que Chociay aponta versões corretas para versos problemáticos, como o seguinte, que também deveria ser heptassílabo:

MAT:2.1.7.1 De Jesus tanto agrado leva

Chociay (74, p. 26) afirma, com base na edição da Academia Brasileira de Letras, que a explicação é trocar *Jesus* por *Deus*. Avaliar quais correções deveriam ser feitas estão fora do escopo desta tese, motivo pelo qual tais casos foram todos marcados como inescandíveis no *corpus*.

**Corpus DUR.** Este é o poema épico *Caramuru*, de Santa Rita Durão (24). Duas edições de referência foram usadas, ambas em formato digital: uma do século XVIII (25) e outra do XIX (26). Não está claro qual é a edição sobre a qual baseou-se a digital. Seguindo o formato da oitava-rima, todas as estrofes contêm exatamente 8 versos. Este é o único *corpus* entre os principais que não possui nenhum verso repetido.

Entre os 33 estrangeirismos (4,9%) marcados no *corpus* estão antropônimos italianos, como *Vespucci* em DUR.6.69.2, e franceses, como *Nicot*, como no seguinte verso:

DUR.7.27.5 Pelo franco Nicot manipulada,

As expansões, em número de 18 (2,7%), são 5 casos de *D.* que são lidos como *Dom* e 13 casos de *S.*, ou seja, *São*.

O grande número de correções (90 ou 13,5%) inclui casos de palavras sem o devido diacrítico, como *delicias* em vez de *delícias* em DUR.3.31.3 e *sabias* em vez de *sábias* em DUR.3.61.6, erros de OCR, como *uni* em vez de *um* em DUR.2.49.4 e *lia* em vez de *na* em DUR.4.9.4, e palavras somente parecidas com o original, como *Recebera* em vez de *Recobra* no seguinte verso:

DUR.2.53.1 Recebera o bom Gupeva um novo alento,

**Corpus cos.** Trata-se do livro conhecido pelo título *Obras*, de Cláudio Manoel da Costa (9). A edição de referência utilizada foi a de 1996 da Nova Aguilar (12), a mesma em que se baseou a digital. Este *corpus* é o único que possui poemas em italiano, contando 488 versos distribuídos em 14 sonetos, 2 cantatas e 3 canções. O tamanho mínimo de estrofe do

<sup>2</sup><http://www.filologia.org.br/pereira/textos/5021-201.htm>

*corpus* é nulo, o que se deve, como explicado anteriormente neste capítulo, aos versos emendados.

Os versos repetidos no *corpus* (92 ou 12,0%) se devem à utilização de estribilhos, especialmente nas canções *Fileno a Nise* e *Nise a Fileno*, que somadas contêm 58 repetições. A maioria das repetições (83 ou 90,2%) aparecem num mesmo poema, mas há também 9 repetições entre poemas distintos, como o seguinte, que aparece também em COS.101.7.1:

COS.108.102.7 Feliz, ó Portugal, feliz mil vezes

As emendas entre versos acontecem, em grande parte, novamente nos estribilhos dos poemas já mencionados. Os dois poemas são paralelos e simétricos, sendo exibidos, na edição impressa, de forma espelhada; para ressaltar a simetria entre palavras parecidas e destacar o estribilho, ocorre a quebra de versos em duas partes, mas eles juntos compõem um hexassílabo, que é justamente a métrica do poema. Ocorrem, portanto, 30 versos emendados nesses dois poemas. Os 8 restantes acontecem em diálogos, em que um verso iniciado por uma personagem é terminado por outra, como nas duas estrofes seguintes, que contêm três versos decassílabos:

COS.108.63.1 Não o sabe cantar qualquer do povo;

COS.108.63.2 Algum somente cortesão polido

COS.108.63.3 E que o canta por lá...

COS.108.64.1 Pois eu o aprovo.

Nenhuma marcação de expansão ou de estrangeirismo foi feita, mas se fez necessária a marcação de uma regularização no *corpus*. O vocábulo *aspid* é utilizado em COS.118.16.1, o que normalmente se tornaria *áspide* numa edição contemporânea. Entretanto, como explica a edição da Nova Aguilar, a vogal final da palavra modernizada quebraria a métrica, dando ao verso 7 sílabas em vez das 6 necessárias:

COS.118.16.1 Qual aspid se afigura

A regularização, portanto, marca a posição do acento tônico desta palavra que não segue as convenções ortográficas do português contemporâneo.

A edição digital desta obra contém uma série de problemas. As correções feitas durante a revisão pontual (89 ou 11,6%) se devem a erros de OCR, como *unia* em vez de *uma* em COS.129.3.5, à ausência de diacríticos, como *magoa* em vez de *mágoa* em COS.52.1.3, a formas parecidas porém não idênticas, como *desprezastes* em vez de *desprezaste* em COS.104.13.3, entre outros. O maior problema, contudo, são os versos faltantes e excessivos, ou seja, versos que a edição digital omitiu ou indevidamente adicionou. Repetem-se os 32 primeiros versos do *Epicédio III* (COS.103) e tre-



chos inteiros na Écloga IV (COS.109.1.1), além das repetições nos poemas em italiano. Faltam também vários versos nas Éclogas III e IV (COS.108.1.1 e COS.109.1.1) e nos poemas em italiano. As contagens de versos faltantes e excedentes na Tabela 4.6 incluem os versos em português e italiano; as correções, não, já que a revisão pontual não foi feita nos versos em italiano.

**Corpus GON.** Este *corpus* contém as *Cartas Chilenas*, de Tomás Antônio Gonzaga (17). A edição de referência adotada foi a de 1996 da Nova Aguilar (12). Esta é uma obra cuja autoria é discutida; deixando tais questões de lado, optou-se por incluir no *corpus* todos os poemas dela, incluindo a *Epístola a Critilo*.

As 30 repetições se devem a versos usados de forma enfática ao longo dos poemas. Do total, 22 aparecem no contexto de um mesmo poema, enquanto há 1 repetição que consta em dois poemas e 7 repetições em três, como o seguinte verso, que aparece duas vezes no poema GON.5 e uma vez nos poemas GON.8 e GON.10:

GON.5.1.1 Maldito, Doroteu, maldito seja

Os 12 versos fragmentários são, na verdade, completamente vazios: constam como pontilhados na edição digital e na impressa.

Os 4 estrangeirismos se devem a 3 palavras em latim, como *manuscristi* em GON.7.10.66 e 1 em inglês, *mylord*, no seguinte verso:

GON.2.7.5 Tem um ar de mylord e a todos trata

As poucas correções que foram feitas se devem sobretudo a palavras sem o devido diacrítico, como *vicio* em vez de *vício* em GON.4.1.62 ou *oficio* em vez de *ofício* no seguinte verso:

GON.13.13.4 C'os moços aprendizes deste ofício?

**Corpus MAG.** Este *corpus* corresponde ao livro *Suspiros Poéticos e Saudades*, de Gonçalves de Magalhães (20). A edição de referência empregada foi a de 1859 (19), que, ao que tudo indica, não foi a mesma sobre a qual se baseou a edição digital. O *corpus* contém um único caso de verso emendado, que deixa parte de um verso em uma estrofe separada, o que explica o tamanho mínimo de estrofe nulo da Tabela 4.3.

Dos 73 versos repetidos (13,0%), a ampla maioria, 69, aparecem dentro de um mesmo poema. Algumas repetições são enfáticas, como o verso seguinte, que aparece também em MAG.9.32.4:

MAG.9.27.4 Santo amor vos inflama.

Há também repetições em estribilhos, que é o caso no poema MAG.23, em

que o seguinte verso aparece 7 vezes:

MAG.23.2.4 Em tudo mentes.

Os estrangeirismos marcados são tipicamente nomes, sejam antropônimos, como *Debret* em MAG.52.15.9, ou topônimos, como *Waterloo* em MAG.39.1.4. A única regularização anotada é a palavra *colibri*, que o ritmo exige seja pronunciada como *colibri* no seguinte verso:

MAG.55.3.5 Onde pousa o colibri, e o gaturamo;

A edição de referência traz a forma *colibre*, sem nenhum diacrítico, o que indica que a pronúncia do poeta era de fato paroxítona e que, a julgar pela grafia, não se trata de uma sístole.

A edição digital desta obra sofreu o menor número de correções entre todos os *corpora* principais: 10 ou 1,8%. São erros de OCR, como *og* em vez de *oh* no seguinte verso:

MAG.4.6.11 Abre-te, og céu azul; deixa minha alma

Há também formas incorretas porém parecidas, como *deveis* em vez de *devéis* em MAG.36.17.3, e modernizações indevidas como *desaparecem* em vez de *desparecem* no verso

MAG.19.13.8 Mas quando se despenham, desaparecem,

**Corpus PIN.** Esta é a tradução de Xavier Pinheiro da *Divina Comédia* de Dante (2). A edição de referência utilizada é a de 1907 (1). A tradução segue a métrica e a disposição do original: são 14.233 decassílabos divididos em 100 poemas, cada um composto de uma sequência de estrofes de 3 versos e uma estrofe final com apenas um.

Há 3 versos repetidos ao longo do *corpus*, dois de forma aparentemente accidental e um intencional. Os accidentais são PIN.15.5.2 e PIN.52.47.2, PIN.77.38.3 e PIN.78.9.2; no original de Dante eles não são idênticos. O terceiro caso, contudo, é um verso fazendo citação a outro anterior, o que acontece também em Dante:

PIN.77.32.3 Onde prospera só quem mal não trama.

PIN.78.47.1 *Onde prospera só quem mal não trama."*

As expansões marcadas no *corpus* (17 ou 1,2%) são *S.* para *São* ou *Santo*, como em PIN.16.34.2 e PIN.21.13.2, ou nomes de letra, como *PP* no seguinte verso:

PIN.43.38.1 Da espada a ponta sete PP me havendo

Os estrangeirismos estão bastante presentes na obra (142 ou 10,0%), e se devem principalmente a expressões latinas, como *Regina coeli* em PIN.90.43.2, e aos antropônimos italianos, como *Bonati* em PIN.20.40.1.

As regularizações se devem a três casos. Dois correspondem aos

nomes *Abraam* e *Adam*, que foram regularizados para *Abraão* e *Adão*, como em PIN.4.19.3 e PIN.30.21.1. A terminação destes nomes em *-am* implicam um acento na penúltima sílaba (como em *cantam*), o que impede a boa escansão de vários versos; não está claro, porém, se estes são erros a serem corrigidos na edição digital, daí as regularizações. O último caso é a palavra *zênite*, que aparece duas vezes no texto. A exemplo do que acontece na obra de Gregório de Matos, esta palavra, para que os dois versos em que aparece sejam escandidos corretamente, precisa ser pronunciada sem a vogal final e de forma oxítona:

PIN.36.1.3 O zênite fica de Solima ao monte.

PIN.96.2.1 Quanto espaço o zênite os equilibra,

As correções se devem à ausência de diacríticos, como *Lavínia* em PIN.4.42.3, palavras excluídas, como *meu* em PIN.10.16.2, e formas incorretas porém parecidas, como *mavórtico* em vez do correto *mavórcio* em PIN.22.2.3. Há 8 versos que foram considerados inescandíveis; trata-se de casos em que pequenos ajustes no texto os tornariam escandíveis, mas que, como estão, ou é impossível escandi-los ou exigiriam a aplicação de regras de acomodação pouco prováveis. Um exemplo é o seguinte, que não pode ser escandido nas 10 sílabas necessárias:

PIN.58.46.2 Aos olhos, dera em fusão, no forno ardente,

**Corpus ANJ.** Esta é a obra completa de Augusto dos Anjos (4). A obra de referência foi a edição de 1994 da Nova Aguilar (3), a mesma que serviu de base para a edição digital.

Os versos repetidos (20 ou 3,0‰) correspondem a 20 versos que aparecem duas vezes cada, todos dentro de um mesmo poema, como ênfase e não estribilho, frequentemente em estrofes inteiras repetidas, como a primeira e a última estrofe da *Barcarola*:

ANJ.54.1.1 Cantam nautas, choram flautas

ANJ.54.1.2 Pelo mar e pelo mar

ANJ.54.1.3 Uma sereia a cantar

ANJ.54.1.4 Vela o Destino dos nautas.

Há 16 versos fragmentários, dos quais 13 completamente vazios; 15 destes versos fragmentários estão no poema *A Meretriz*, ANJ.67, e 1 no poema *M. B.*, ANJ.261. Os 2 versos emendados, ANJ.116.4.3 e ANJ.214.1.36, são, na verdade, emendas entre um verso que começa normalmente mas termina na assinatura do poema.

Este *corpus* contém o maior número relativo de estrangeirismos e de expansão entre todos os *corpora* principais: são 26 expansões (3,9‰) e

84 estrangeirismos (12,7%). As expansões são pronomes de tratamento, como *D.* para *dona* em ANJ.281.6.9 e *V.* para *você* em ANJ.280.4.3, nomes de letra, como *Raio X* em ANJ.1.11.2, e números, como *1\$200* em ANJ.01:25-86-4 e 33 em ANJ.31.1.3.

Ao todo 21 correções foram feitas à edição digital, que se devem principalmente a erros de OCR, como um zero em vez da letra *O* no verso ANJ.284.12.4, à ausência de diacríticos em palavras comuns, como em *va-rio* em vez de *vário* no verso ANJ.149.2.4, e à ausência de diacríticos em vocábulos que o autor requer sejam sempre pronunciados com sístole, como *elégia* em ANJ.1.30.1 e *periféria* no verso

ANJ.25.104.4 Para alcançar depois a periféria!

A grafia com diacrítico é atestada na edição de referência.

**Corpus FIG.** Este é o livro *Poesias*, de Araújo Figueredo (6). A edição de referência utilizada foi a de 1966 da Academia Catarinense de Letras (5), a mesma em que se baseou a digital.

Os versos repetidos neste *corpus*, em número de 131 (15,9%), têm várias origens. Podem aparecer no mesmo poema, como em FIG:2.6.2.1; em poemas distintos, como os versos FIG:2.3.51.2 e FIG:02:23-29-2; podem estar em estrofes idênticas, como em FIG:2.4.1 e FIG:2.4.10; podem ser invocações que lembram estribilhos, como em FIG:2.26.15.2; ou podem aparecer diversas vezes para indicar os falantes em diálogos, como em FIG:2.28.2.10.

Há apenas uma expansão anotada no *corpus*. Trata-se da expressão *A-B-C* no seguinte verso:

FIG:3.31.1.2 Sabe de cor a carta do A-B-C.

As diversas correções feitas à edição digital (39 ou 4,7%) correspondem a erros de OCR, como *unia* em vez de *uma* em FIG:1.45.1.2, palavras incorretas, como *Doa* em vez de *Dos* em FIG:2.3.12.2, versos indevidamente juntados em um só, como em FIG:2.26.1.4 e FIG:2.26.1.5, e sequências suprimidas, como *gelo*, e *às vezes*, em FIG:4.40.3.3.

Este é o *corpus* com o maior número relativo de versos considerados inescandíveis (33 ou 4,0%). Estes versos, em sua maior parte, poderiam ser levemente alterados para que pudessem ser escandidos corretamente. O seguinte verso deveria conter 12 sílabas, mas possui no mínimo 13:

FIG:1.132.2.4 Uivam nas praias os cães e mugem bois, errantes...

Se lhe fosse retirado o artigo *os* antes de *cães*, o verso se tornaria um alexandrino. Da maneira como está, contudo, parece ser impossível de escandir o verso. Há também casos de poemas polimétricos em que a separação

das palavras em versos parece ter sido feita de forma incorreta, como na estrofe seguinte:

FIG:2.26.22.1 — Àque-la que se lembra do jumento e da

FIG:2.26.22.2 meiga vaquinha, que se encheram de eterno sentimento,

FIG:2.26.22.3 na hora em que nasceu Jesus, o amado

FIG:2.26.22.4 filho do seu amor, seja o nome louvado

FIG:2.26.22.5 no maior esplendor!

O segundo verso conta com no mínimo 15 sílabas, sem nenhuma possibilidade de ser reduzido a 12 ou menos. Uma reorganização drástica das fronteiras dos versos, sem contudo alterar-lhes a ordem das palavras, resultaria nesta possível escansão:

1	2	3	4	5	6	7	8	9	10	11
— À-	que-	la	que	se	lem-	bra	do	ju-	men-	to
e	da	mei-	ga	va-	qui-	nha,	que	se en-	che-	ram
de e-	ter-	no	sen-	ti-	men-	to,				
na	ho-	ra em	que	na-	sceu	Je-	su-	s, o a-	ma-	do
fi-	lho	do	seu	a-	mor,					
se-	ja o	no-	me	lou-	va-	do				
no	mai-	o-	res-	plen-	dor!					

Com essa escansão, tem-se três decassílabos heroicos e três hexassílabos, além de seis versos que rimam, em vez dos dois no original.

**Corpus SIL.** Esta é a obra completa de Delminda Silveira, que, na Literatura Digital, foi dividida em vários documentos (27, 28, 29, 30, 31). A edição de referência utilizada para conferência foi a mesma em que se baseou a edição digital: a de 2009 da Academia Catarinense de Letras (32).

O livro divide-se em partes. Algumas delas correspondem a outros livros previamente publicados, enquanto outras coletam material disperso em veículos como jornais e revistas. Por este motivo, existem poemas que figuram mais de uma vez na obra, o que explica a grande quantidade de versos repetidos, 1.286 (111,4%). Destes, 148 aparecem no contexto de um mesmo poema, enquanto 944 estão em dois poemas e 170 estão em três.

As expansões marcadas no *corpus* incluem *D.* para *Dom*, como em SIL:5.126.47.4, e *S.* para *São*, como em SIL:5.126.47.4. Os estrangeirismos anotados são antropônimos, como *Byron* em SIL:5.42.1.1 e *Dewet* em SIL:5.21.1.1, e também substantivos, como *bouquet* em SIL:2.34.2.2.

O grande número de regularizações anotadas, 55 ou 4,3%, se deve à presença de 54 casos em que a conjunção *se* está escrita com a forma an-

tiga *si* (a forma oblíqua tônica *si* não precisa e não foi regularizada). Não está claro por que foi tomada a decisão de conservar a grafia antiga, mas, como se trata de uma questão que interfere na escansão dos versos, todas as instâncias encontradas foram regularizadas. Além destes 54 casos, há ainda a palavra *cactus* no verso seguinte, que consta sem o diacrítico na edição impressa e na digital:

SIL:1.19.3.2 a branca flor do cactus, formosa;

Como ocorre também a palavra *cactos* no verso SIL:1.100.4.2 e não se conhece a intenção dos organizadores da edição, optou-se por regularizar a forma sem corrigi-la.

A presença de poemas duplicados e triplicados foi explorada para corrigir versos em que a revisão pontual encontrou problemas. Por exemplo, o verso SIL:5.127.4.3 na edição impressa e na digital não pode ser escandido nas 12 sílabas exigidas pelo metro:

SIL:5.127.4.3 que lágrimas fel tão d'alma derramaste

Por fazer parte de um poema duplicado, *Na Soledade*, pode-se esclarecer que está faltando a preposição *de*, presente no verso correspondente da duplicata:

SIL:1.78.4.3 — que lágrimas de fel tão d'alma derramaste

A edição impressa, que a digital seguiu, contém alguns erros de digitação. No soneto SIL:5.106, a palavra *sondas* foi transporta do último verso do segundo quarteto, SIL:5.106.2.4, para o último verso do primeiro terceto, SIL:5.106.3.3, o que deixou os dois versos envolvidos com problemas métricos e quebrou o esquema rimático. Mais uma vez, por existir uma outra versão do mesmo poema, SIL:1.68, os versos puderam ser corrigidos. Também uma instrução de formatação foi deixada no meio de um verso. Ao fim do poema *A Sós Fala Don Juan*, SIL:5.166, o último verso está assim na edição impressa (e na digital):

SIL:5.165.10.2 — onde estará, porém, a não ital (o resto sim) única mulher

O trecho *não ital (o resto sim)* evidentemente não faz parte do verso.

Além dos erros presentes na edição impressa, a edição digital introduziu mais alguns, principalmente de OCR, como a sequência *n'* (a letra *N* mais apóstrofo), que com bastante frequência aparece como *ri* na edição digital. Existem ainda 20 versos que constam na edição digital mas não na impressa: há uma estrofe supérflua ao fim do poema *Santa Catarina*, SIL:5.7, e o poema *Esperança*, SIL:5.159, aparece duas vezes em sequência na edição digital.

Tabela 4.7.: Os *corpora* poéticos secundários, em ordem cronológica segundo o ano de morte do autor. Os códigos entre parênteses indicam itens já presentes nos *corpora* primários.

Código	Autor	Morte	Obra
(CAM)	Luís de Camões	1580	Os Lusíadas
(DUR)	Santa Rita Durão	1784	Caramuru
MAN	Cláudio Manoel da Costa	1789	Vila Rica
GAM	Basílio da Gama	1795	O Uruguai
DIA	Gonçalves Dias	1864	Os Timbiras
OLI	Alberto de Oliveira	1937	Margens de Ouro e de Esmeralda
LIM	Jorge de Lima	1953	Permanência de Inês

#### 4.3. *Corpora* poéticos secundários

Os *corpora* poéticos secundários são utilizados num dos casos de uso do Capítulo 6, em que a proporção de decassílabos heroicos e sáficos são comparados entre os *corpora*. Por não terem a intenção de validar o sistema, não estão aqui descritos em tantos detalhes quanto os primários. Eles foram estabelecidos da mesma maneira que os primários. Dois dos *corpora* poéticos secundários coincidem, entretanto, com os primários e portanto utilizam o mesmo código de três letras e não serão aqui descritos novamente: trata-se de CAM e DUR. A Tabela 4.7 mostra as obras que compõem os *corpora* secundários, enquanto a Tabela 4.8 exhibe as quantidades de versos, estrofes e poemas que os compõem.

Os erros corrigidos constam da Tabela 4.9. É importante observar que dois dos *corpora*, OLI e LIM, foram derivados de edições impressas, o que faz com que a contagem de versos corrigidos, faltantes e excedentes perca a razão de ser. Essas contagens refletem melhorias feitas nas edições digitais que servem de base para os *corpora*. Ora, se o próprio autor da tese responsabilizou-se pela produção das edições digitais desses dois *corpora*, não faz sentido que ele mesmo tente melhorá-las. Dito de outra forma, quaisquer erros encontrados durante o estabelecimento das edições digitais foram nelas incorporados e portanto a contagem de erros a partir das edições digitais finais é por definição nula.

A partir de cada um dos *corpora* secundários é derivado outro contendo apenas os 100 primeiros versos. Estes *corpora* reduzidos são denotados pelo mesmo código de três letras mais o sufixo 100; assim, o *corpus*

Tabela 4.8.: Contagens de versos, estrofes e poemas dos *corpora* poéticos secundários.

<i>Corpus</i>	Versos		Estrofes			Poemas			
	Total	Total	Tamanho			Total	Tamanho		
			Mín.	Máx.	Méd.		Mín.	Máx.	Méd.
(CAM)	8.816	1.102	8	8	8,0	10	696	1.248	881,6
(DUR)	6.672	834	8	8	8,0	10	592	736	667,2
MAN	2.718	160	3	88	17,0	10	178	458	271,8
GAM	1.377	6	11	365	229,5	5	150	365	275,4
DIA	2.032	151	2	60	13,5	5	60	609	406,4
OLI	184	11	3	39	16,7	1	184	184	184,0
LIM	143	18	7	8	7,9	1	143	143	143,0
	21.942	2.282	2	365	9,6	42	60	1.248	522,4

Tabela 4.9.: Número de versos corrigidos, faltantes, excedentes e inescandíveis nos *corpora* poéticos secundários. O número de versos corrigidos, faltantes e excedentes de OLI e LIM não são indicados porque estes *corpora* não tiveram como base edições digitais, o que impede essas contagens.

<i>Corpus</i>	Corrigidos		Faltantes		Excedentes		Inescandíveis	
	Total	‰	Total	‰	Total	‰	Total	‰
(CAM)	117	13,3	3	0,3	3	0,3		
(DUR)	90	13,5						
MAN	17	6,3						
GAM	2	1,5			1	0,7		
DIA	63	31,0	37	18,2				
OLI	-	-	-	-	-	-		
LIM	-	-	-	-	-	-		
Total	289	13,2	40	1,8	4	0,2		



OLI100 é constituído dos 100 primeiros versos do OLI. Esta derivação é necessária devido à natureza do estudo de caso já referido.

A seguir segue uma breve descrição dos *corpora* secundários, exceto CAM e DUR.

**Corpus MAN.** Este é o poema épico *Vila Rica*, de Cláudio Manuel da Costa. Para constituir o *corpus* foi utilizada a edição digital da Literatura Digital (11). A edição de referência empregada foi a de 1839 (10). As correções feitas se devem principalmente a erros oriundos da etapa de OCR da digitalização, como a palavra *oficio* em vez do correto *ofício* no verso MAN.4.4.21.

**Corpus GAM.** Trata-se do épico *O Uruguai*, de Basílio da Gama. Novamente a edição da Literatura Digital foi utilizada (16). A edição de referência adotada foi a de 1769 (15). As duas correções feitas foram remoções de um -s final que impedia a boa escansão de versos, como em *dar-lhes* em vez de *dar-lhe* em GAM.2.1.70.

**Corpus DIA.** Este é o épico *Os Timbiras*, de Gonçalves Dias. A edição da Literatura Digital (14) foi utilizada. Como edição de referência empregou-se a de 1857 (13). O grande número de correções se deve, sobretudo, à grafia *Itajubá* em vez do esperado *Itajuba*. Foram também detectados 37 versos ausentes da edição digital, espalhados em várias partes da obra.

**Corpus OLI.** Este é o poema *Margens de Ouro e de Esmeralda*, de Alberto de Oliveira. O *corpus* foi criado diretamente a partir da edição de 1928 (23). A edição impressa, em meio físico, foi convertida em meio digital pelo próprio autor.

**Corpus LIM.** Este é o Capítulo IX, intitulado *Permanência de Inês*, do livro *Invenção de Orfeu*, de Jorge de Lima. De forma similar ao OLI, este *corpus* deriva-se da edição impressa, neste caso de 1952 (18).



## 5. O SISTEMA AOIDOS

1	2	3	4	5	6	7	8	9	10	11	12
Pa-	re-	ce in-	crí-	vel	que o	fa-	zer	po-	é-	ti-	co
Es-	te-	ja	pre-	so a	re-	gras	tão	for-	mais!		
A-	le-	xan-	dri-	no e be-	roi-	co	são	nor-	mais,		
Ma-	s es-	te	sá-	fi-	co é	bem	mai-	s ber-	mé-	ti-	co.

Glauco Mattoso

Este capítulo descreve em detalhes o sistema Aoidos. As etapas do sistema encontram-se exemplificadas na Figura 5.1. O sistema recebe como entrada poemas codificados em XML (1) e, ao fim, produz uma escansão para cada verso (9). O poema é a maior unidade de análise: os versos de um poema não tem influência sobre os de outro.

O sistema começa com quatro etapas que trabalham nas palavras individuais dos versos. O pré-processamento normaliza alguns aspectos do texto e extrai as palavras (2). O sistema, então, para cada palavra, encontra então a vogal tônica (3) e faz a divisão silábica (4). Com base nas etapas anteriores, uma transcrição fonética é realizada (5), encerrando o processamento do sistema com base em vocábulos isolados.

O sistema concatena a transcrição fonética das palavras dos versos e, a partir dessa transcrição inicial, inúmeras variantes são produzidas para cada verso (6), cada uma com diferentes combinações de metaplasmos aplicados. Na figura, o número que precede cada transcrição é o número de sílabas poéticas da variante, e a sequência de números que a sucede é seu esquema rítmico. As sílabas sublinhadas indicam diferenças com relação à primeira variante.

Levando em consideração todas as variantes de todos os versos de um poema, o sistema tenta descobrir qual é a métrica que os versos seguem (7). O resultado dessa tentativa pode ser *poema isométrico decassilábico* ou *poema com eneassílabos e tetrassílabos*. Na figura, o sistema conclui que o poema é isométrico, com dez sílabas por verso. As variantes são então filtradas de acordo com a métrica: se o poema foi classificado como decassílabo, como na figura, então somente as variantes com 10 sílabas são consideradas. Por fim, o sistema busca, entre as variantes sobreviventes,

1	Entrada	<l n="2">Nunca lhe falta um Gil, que o persiga,</l>										
2	Pré-processamento	nunca	lhe	falta	um	gil	que	o	persiga			
3	Vogal tônica	nunca	lhe	falta	um	gil	que	o	persiga			
4	Divisão silábica	nun-ca	lhe	fal-ta	um	gil	que	o	per-si-ga			
5	Transcrição fonética	/'nũ.ka/ /ɫɪ/ /'fal.ta/ /ũ/ /'zil/ /kɪ/ /ø/ /per'si.ga/										
6	Geração de elocuições	11	/'nũ.ka.ɫɪ'fal.ta.ũ'zil.kɪ.ø.per'si.ga/								1-4-7-11	
		10	/'nũ.ka.ɫɪ'fal.ta.ũ'zil.kjɔ.per'si.ga/								1-4-7-10	
		10	/'nũ.ka.ɫɪ'fal.tũ'zil.kɪ.ø.per'si.ga/								1-4-6-10	
		9	/'nũ.ka.ɫɪ'fal.tũ'zil.kjɔ.per'si.ga/								1-4-6-9	
7	Análise métrica	8 9 10 11 12 13										
8	Análise rítmica	1-4-7-10 1-4-6-10										
9	Saída	1	2	3	4	5	6	7	8	9	10	11
		Nun-	ca	lhe	fal-	ta	um	Gil,	que	o	per-si-	ga,
		'nũ	ka	ɫɪ	'fal	tũ	'zil	kɪ	u	per	'si	ga

Figura 5.1.: Etapas do sistema e respectivos exemplos.

aquelas com uma melhor adequação rítmica (8). O resultado produzido pelo sistema é uma escansão (9), que pode ser exibida tanto de forma fonética quanto mapeada sobre os símbolos do texto original.

Todas as etapas do sistema são descritas em detalhe a seguir, cada uma em sua própria seção. Por fim, alguns detalhes técnicos são mencionados e algumas considerações metodológicas apresentadas.

5.1. Entrada

**Codificação.** O formato TEI, baseado em XML, é usado pela comunidade de humanidades digitais para codificar obras textuais. O objetivo deste formato é codificar a semântica presente no documento; as possíveis questões de formatação são deixadas para uma folha de estilo. Os elementos definidos pelo TEI cobrem um amplo terreno: existem elementos específicos para a codificação de dicionários, manuscritos, edições críticas, transcrição de fala, etc. De interesse à presente tese é o Capítulo 6 das atuais diretrizes publicadas pelo consórcio TEI (167), que descreve os elementos para a codificação de formas em verso.

Um verso é codificado em TEI através do elemento <l> (do inglês *line*). Um agrupamento de versos, ou seja, uma estrofe, é representado pelo

```

<div type="poem">
  <head>O MORCEGO</head>
  <lg>
    <l>Meia-noite. Ao meu quarto me recolho.</l>
    <l>Meu Deus! E este morcego! E, agora, vede:</l>
    <l>Na bruta ardência orgânica da sede,</l>
    <l>Morde-me a goela ígneo e escaldante molho.</l>
  </lg>
  <lg>
    <l>"Vou mandar levantar outra parede..."</l>
    <l>— Digo. Ergo-me a tremer. Fecho o ferrolho</l>
    <l>E olho o teto. E vejo-o ainda, igual a um olho,</l>
    <l>Circularmente sobre a minha rede!</l>
  </lg>
  <lg>
    <l>Pego de um pau. Esforços faço. Chego</l>
    <l>A tocá-lo. Minh'alma se concentra.</l>
    <l>Que ventre produziu tão feio parto?!</l>
  </lg>
  <lg>
    <l>A Consciência Humana é este morcego!</l>
    <l>Por mais que a gente faça, à noite, ele entra</l>
    <l>Imperceptivelmente em nosso quarto!</l>
  </lg>
</div>

```

Figura 5.2.: Exemplo do formato TEI: o poema ANJ.3.

elemento <lg> (do inglês *line group*). Outras divisões de um poema ou de uma obra poética são definidas com o elemento genérico <div>; a função específica da divisão pode ser indicada através do seu atributo *type*. A Figura 5.2 mostra um poema inteiro codificado neste formato. Existem atributos do TEI que podem, de forma opcional, ser usados para especificar, por exemplo, a métrica de um verso ou poema (atributo *met*). Tais atributos não são utilizados pelo Aoidos.

A intenção de utilizar o formato TEI é permitir que o mesmo arquivo sirva tanto para o Aoidos quanto para outros sistemas. Assim, o trabalho de codificação não é feito somente para um sistema. Em particular, a utilização do TEI permite que o arquivo seja processado por uma folha de estilo para geração de uma edição em HTML ou PDF.

Ao receber um arquivo XML, o Aoidos busca as estrofes no documento. São ignoradas aquelas que se encontram dentro do elemento <open>, que é utilizado para codificar epígrafes. Todas as estrofes que se en-

```

<l>
  Custa
  <choice>
    <orig>1$200</orig>
    <seg type="exp">mil e duzentos</seg>
  </choice>
  ao lojista!
</l>

```

Figura 5.3.: Exemplo de verso com números (ANJ.25.86.4).

contram na mesma divisão do documento são consideradas parte de um único poema. O sistema utiliza, portanto, uma hierarquia simples: um poema é um conjunto de estrofes, e uma estrofe é um conjunto de versos.

**O elemento choice.** Alguns casos especiais devem ser aqui mencionados. Como explicitado na introdução, o Aoidos parte de um verso escrito com palavras em português. Para que versos contendo números, siglas ou estrangeirismos sejam processados com sucesso pelo sistema, o elemento <choice> do TEI pode ser usado. Tomemos o verso da Figura 5.3. Ele contém um preço, mil e duzentos réis, escrito na notação da época: *1\$200*. O elemento <orig>, inserido dentro do <choice>, contém este texto original. Já o <seg>, também dentro de <choice>, contém a expansão daquele valor: *mil e duzentos*. O fato de que a expansão não contém o nome da moeda aponta para a grande dificuldade em tratar tais casos de forma automática: ao ler o verso sem considerações métricas em mente não é possível saber se a moeda deve ou não ser pronunciada.

Além da expansão de números e siglas, outro caso importante é o uso de palavras estrangeiras nos versos. O verso da Figura 5.4 contém a expressão latina *ignis sapiens* que, se lida de acordo com as regras do português, teria o acento nas vogais erradas: *ignís sapiéns*. O elemento <orig>, neste caso, contém a expressão original marcada ainda com o elemento <foreign>, que indica um fragmento em outra língua. O atributo `xml:lang` desse último elemento é usado para indicar em qual língua estrangeira o fragmento está escrito. Tais marcações são destinadas à folha de estilo, que pode escolher formatar o fragmento em itálico ou adicionar uma ligação para uma ferramenta de tradução. O elemento <seg>, com o atributo `type="apo"` (aportuguesamento), contém o texto a ser lido especificamente pelo Aoidos. O texto do aportuguesamento deve estar escrito de acordo com as regras do português.

```

<l>
  Ou, pelo menos, o
  <choice>
    <orig>
      <foreign xml:lang="la">ignis sapiens</foreign>
    </orig>
    <seg type="apo">ígnis sápiens</seg>
  </choice>
  do Orco
</l>

```

Figura 5.4.: Exemplo de verso com palavras estrangeiras (ANJ.12.15.2).

A transcrição de palavras estrangeiras e a expansão de números é importante para a análise cuidadosa de um *corpus*, mas não é obrigatória; o Aoidos tentará de qualquer forma processar todos os versos que encontrar, ainda que não vá fazer tentativas automáticas de aportuguesamento ou de expansões.

**Versos emendados.** Uma terceira e última estipulação concerne à disposição gráfica de alguns versos no poema original. Diferente dos dois casos descritos há pouco, este não tem o objetivo de contornar uma limitação do sistema, mas sim de manter compatibilidade com outros sistemas que venham a processar o mesmo arquivo. Consideremos o fragmento da Figura 5.5, que mostra a parte final de um poema. Além das duas últimas estrofes do soneto, a figura mostra a assinatura (<signed>) e a data (<dateline>) do poema, ambas contidas no elemento de fechamento <closer>. O último verso do poema não termina no fim da linha definida pelo elemento <l>, mas sim na assinatura do poema. Com efeito, se desconsiderarmos a assinatura do poema, o último verso fica com apenas 8 sílabas poéticas ao invés das 10 que os outros versos pedem, e a rima do último verso da estrofe anterior é perdida. A solução adotada para este caso é encadear os elementos que compõem um mesmo verso: o primeiro elemento (necessariamente um <l>) aponta para o próximo fragmento através do atributo TEI next; os próximos elementos (quaisquer) possuem o identificador, xml:id, ao qual se refere o verso anterior e podem também apontar para o próximo elemento que ajuda a compor o verso. No caso de dois ou mais elementos <l> serem encadeados, eles são considerados apenas como integrantes da sequência e não são processados de forma independente.

```

<lg>
  <l>Somente assim festejarei teus anos;</l>
  <l>Enquanto outros que podem, dão-te enganos,</l>
  <l>Jóias, bonecos de formoso busto,</l>
</lg>
<lg>
  <l>Eu só encontro no primor de rima</l>
  <l>A justa oferta, a jóia que te exprima</l>
  <l next="#mano-augusto">0 amor fraterno do teu mano</l>
</lg>
<closer>
  <signed xml:id="mano-augusto">Augusto.</signed>
  <dateline>Em 28 de abril de 1901</dateline>
</closer>

```

Figura 5.5.: Exemplo de versos emendados (fim do poema ANJ.1.16).

**Discussão.** O fato de o Aoidos depender de marcações manuais para casos como palavras estrangeiras e numerais constitui uma limitação do sistema. Com o objetivo de convencer o leitor de que não se trata de uma questão trivial, são apresentados a seguir alguns problemas a serem enfrentados para que a marcação manual não seja necessária.

A pronúncia de numerais escritos com algarismos nos poemas enfrenta alguns desafios, ainda que existem casos fáceis como 21 (ANJ.55.24.3) e 6ª (ANJ.55.7.3), que devem ser lidos como *vinte e um* e *sexta*. Alguns numerais podem ser mais difíceis por flexionarem em gênero; o numeral 300 no verso ANJ.55.3.2 deve ser lido no masculino, *trezentos*, porque se refere à palavra *quilos*, mas poderia muito bem ser pronunciado *trezentas* se se referisse a *pessoas*. Valores monetários também ocorrem em poemas; quando o leitor encontra 1\$200 (ANJ.25.86.4), pode não ser evidente se o nome da moeda, *réis* neste caso, deve ser pronunciado também. Datas apresentam a particularidade de que os números dos dias devem ser pronunciados usando numerais cardinais, exceto quando é o primeiro dia do mês, caso em que o numeral ordinal deve ser utilizado. No poema ANJ.50, o poeta fala de um evento que acontece *a 1 de janeiro*. A escolha natural seria, portanto, pronunciar este trecho como *a primeiro de janeiro*; entretanto, ao fazê-lo, o leitor estaria desrespeitando a métrica do verso. Para obter as 10 sílabas métricas do verso, é preciso utilizar a leitura *a um de janeiro*, que é pouco intuitiva, pelo menos de acordo com os parâmetros atuais. Dificuldade ainda maior apresentam os numerais romanos. Eles podem ser confundidos com palavras legítimas da língua, como é o



caso de VI. Ademais, não é feita distinção na escrita entre ordinais e cardinais; *X* pode ser tanto *dez* quando *décimo*.

A expansão de abreviaturas não pode ser feita apenas com um dicionário. Há casos em que a mesma abreviatura tem expansões diferentes; é o que acontece com a abreviatura *D.*, que Augusto dos Anjos usou para se referir a *Dona* e Camões para se referir a *Dom*. Determinar qual é a expansão correta depende não apenas de uma análise sintática, para saber que *um D. Fuas* (CAM.1.12.3) só pode ser masculino, mas também de um mecanismo para classificar nomes em masculino e feminino, como é o caso de um verso como o seguinte:

ANJ.281.6.9 Bravos de D. Donzinha

Palavras estrangeiras apresentam um grande desafio. Primeiro, é preciso saber se uma determinada palavra é estrangeira ou não. Uma palavra como *kinnors* (ANJ.27.8.2) é evidentemente estrangeira, mas não é tão óbvio que *papirus* (ANJ.1.3.5) o seja, e mais difícil ainda é saber que *sítio* (PIN.56.2.3) é uma palavra latina (pronunciada, portanto, *sítio*) e não a 1ª pessoa do singular do presente indicativo do verbo *sítiar*. Em segundo lugar, deve-se identificar a língua estrangeira em que a palavra está escrita. Pode parecer claro que *argots* (ANJ.25.77.1) é francesa e *iceberg* (ANJ.98.23.4) é inglesa, mas há casos mais difíceis, como o nome próprio *Sizi* (PIN.83.36.3), que é italiano. Em terceiro lugar, se sabemos que uma palavra é estrangeira e que ela pertence de fato a uma língua específica, é necessário saber como ela deve ser pronunciada na língua estrangeira para então obter uma forma de aportuguesar a sua pronúncia. Em alguns casos, talvez se poderia utilizar uma pronúncia pouco precisa, mas há casos em que a palavra estrangeira precisa rimar com uma escrita em português. É o caso da referida palavra inglesa *iceberg*, que Augusto dos Anjos faz rimar com *se ergue*; a pronúncia perfeita em inglês destruiria a rima. Também não se trata de ler a palavra como se fosse português, quando é preciso rimar *Grasset* com *Fé* (ANJ.251.1.44); aqui o *t* da primeira palavra precisa ser ignorado. A atitude oposta, a de tentar reproduzir a pronúncia da língua estrangeira da maneira mais fiel possível em português também falha: Augusto dos Anjos faz *Vinci* (ou seja, Leonardo da Vinci) rimar com *lince* (ANJ.12.29.2). Seria possível uma pronúncia em português mais próxima do original: poderíamos pronunciar o nome como se estivesse escrito *Vínchi* ou *Víntchi*. Ainda seria necessário encontrar uma maneira de distinguir onde termina a língua estrangeira e o onde começa a vernácula em palavras como *clownesco* e *byroniana* (ANJ.281.1.2, ANJ.247.1.1).

Por fim, encontram-se ainda nomes de letras nos versos, escri-

tos não por extenso mas sim com o próprio grafema, ou seja, escritos *X* (ANJ.1.1.1.2) e não *xis*. Como os nomes das letras não são escritos necessariamente em maiúsculas, é possível que certos nomes sejam confundidos com o artigo definido feminino *a*, com a conjunção *e* e com o artigo definido masculino *o*. Novamente, apenas uma boa análise sintática seria capaz de determinar qual é o caso.

Essas limitações todas são extremamente importantes pois produzirão erros quando um *corpus* não tratado for processado. Não houve tentativa de tratá-las neste trabalho por dois motivos: primeiramente porque elas não surgem de problemas de *escansão* propriamente dita; segunda-mente porque, como espera-se estar claro após a discussão acima, são problemas difíceis de serem resolvidos e que merecem um esforço de pesquisa separado. A resolução destas limitações passará necessariamente pela incorporação ao sistema de uma análise sintática e de um mecanismo de interação entre as partes iniciais e finais do algoritmo. Esse último requisito se deve ao fato de que vários casos só podem ser resolvidos a contento quando a métrica final é avaliada (caso do *um de janeiro*) ou quando as rimas são comparadas (caso do *iceberg*).

## 5.2. Pré-processamento

O sistema faz algumas alterações no texto lido do arquivo XML e então separa o texto em palavras. As alterações são poucas e tentam dar uma certa uniformidade ao texto no que se refere ao uso de apóstrofes.

**O apóstrofo.** O apóstrofo não é utilizado na grafia da prosa em português da maneira disseminada como ele o é, por exemplo, em inglês (*don't*) e francês (*l'âge*). Utilizamos-lo sobretudo em expressões fixas, como em *pau-d'arco* e *pau-d'água*. Na poesia, porém, o apóstrofo vê larga utilização: é uma das principais ferramentas gráficas que o poeta ou seu editor utilizam para indicar ao leitor a presença de certos fenômenos de acomodação. Deste modo, se o poeta escreve *felicidade*, ele não deixa dúvidas sobre a síncope de uma vogal que via de regra é bem ouvida na pronúncia corriqueira da palavra *felicidade* no Brasil. A utilização do apóstrofo, entretanto, não é nem obrigatória, nem padronizada. A obra de um mesmo poeta pode ou não conter apóstrofes conforme o livro consultado. Portanto, o sistema pode buscar compreender para sua própria vantagem qual foi a intenção do poeta ou do editor ao utilizar o apóstrofo, mas não pode depender dele para escandir os versos.

O que o Aoidos faz, em geral, é simplesmente remover o apóstrofo do texto. Quando é utilizado no meio da palavra, o resultado normalmente é uma palavra grafada de forma se não correta, ao menos processável pelo sistema: de *p'ra* obtém-se *pra*; de *másc'ra*, *máscra* (com diacrítico redundante); de *pét'las*, *pétlas* (novamente com diacrítico supérfluo). Um maior desafio são os apóstrofos que reúnem duas ou mais palavras, como *minh'álma*, *d'amor* e *d'um'álma*. Também em casos assim o sistema remove o apóstrofo, porém desta vez tomando o cuidado de remover um eventual *h* do início da segunda palavra, como em *d'honras*: assim, ele transforma *d'homens* em *domens*, *d'amargores* em *damargores*, *d'um'álma* em *du-malma*. Em caso de junção de duas palavras acentuadas, como *minh'álma*, a remoção do apóstrofo efetivamente remove um acento.

Um caso bastante particular é o da preposição *com*. Ela frequentemente sofre eclipse, isto é, a perda de sua trava nasal, e uma consequente sinalefa com a vogal que a sucede. É muito comum, por exemplo, que ela se funda às formas do artigo definido, de modo que a sequência *com as* seja pronunciada numa sílaba só. É um fenômeno muito comum na fala descuidada contemporânea do português brasileiro e tem sido amplamente utilizado na poesia pelo menos desde Camões. Nos textos poéticos em que o autor tem a intenção de ajudar o leitor indicando graficamente esta eclipse, é muito comum que ela seja representada assim: *co'as*, *co'ò*. Contudo, se o sistema simplesmente remover o apóstrofo, serão obtidas formas como *coas*, *coo*, que serão mais tarde necessariamente interpretadas como tônicas. De fato, tais formas até mesmo coincidem com flexões do verbo *coar*, evidentemente tônicas. Portanto, dado que de qualquer forma o sistema deve estar preparado para fazer eclipses não indicadas graficamente, o pré-processamento *desfaz* a marcação delas no texto, transformando *co'as* em *com as*, *co'ò* em *com o*, e assim por diante. O sistema, ao adotar este comportamento, está assumindo o risco de que posteriormente conseguirá detectar de qualquer modo as eclipses; se não conseguir fazê-lo, escandirá o verso incorretamente.

**Extração das palavras.** O texto dos versos é dividido em palavras extraíndo-se sequências contíguas de caracteres reconhecíveis como grafemas; os demais caracteres são desprezados. Assim, sinais de pontuação, números e outras marcações são ignoradas pelo sistema. Em particular, o hífen não é reconhecido como parte das palavras, o que merece uma justificativa. O hífen é usado em português para separar os elementos de uma palavra composta: *cavalo-vapor*. Este uso do hífen reflete a intenção de ex-

pressar que o significado da palavra composta não é diretamente derivado de suas partes (não se trata de um cavalo transformado em vapor ou algo do gênero). Com base nesta observação, o Aoidos extrai os componentes de uma palavra composta e os processa individualmente. Esta abordagem apresenta certos problemas, pois o hífen não é usado exclusivamente em palavras compostas por outras que possuem uma vida acentual independente: o hífen é utilizado também para evitar conflitos ortográficos indesejáveis entre um prefixo e o elemento que o segue. Por exemplo, escreve-se *anti-humano* e assim preserva-se o aspecto da palavra *humano*, que se perderia se o composto fosse escrito *\*antiumano*. A ortografia não marca, contudo, o acento do prefixo *anti*: se o prefixo for lido de forma independente, será acentuado na segunda sílaba. A ortografia vigente prescreve diacríticos somente para os prefixos *pré*, *pós* e *pró*. Os prefixos sem diacrítico seguidos de hífen são tratados como átonos pelo sistema, independentemente de sua pronúncia na prática; assim o prefixo em *anti-horário* e *antiofídico* recebem o mesmo tratamento. A exceção são os prefixos que se escrevem da mesma forma que palavras tônicas, como *contra* e *entre*, que recebem o tratamento regular.

**Remoção da pontuação.** A remoção peremptória da pontuação pode causar certo espanto. De fato, a pontuação pareceria ser uma maneira de o poeta indicar pausas na leitura. Bandeira (42), por exemplo, indignou-se com o comentário de Alberto de Oliveira de que o seguinte verso de Gonçalves Dias estaria, “se não errado, frouxo”:

DIA. 3.22.3 Grita; e os seus, medrosos, receiando

Um verso *frouxo* é aquele que, ao serem feitas todas as devidas sinalefas, possui um número de sílabas menor do que o esperado. A frouxidão do verso, denunciada por Alberto de Oliveira, dever-se-ia à sinalefa entre *Grita* e *e os*. Manuel Bandeira argumenta que, na verdade, pode-se sentir uma pausa após *Grita*. O ponto e vírgula, poder-se-ia argumentar, é que a está indicando. Proença (154) chega mesmo a atribuir uma unidade silábica a uma vírgula para fazer o seguinte verso de Augusto Meyer ser heptassilábico:

Vida minha, tem pena,

Com efeito, o único modo de contar 7 sílabas nesse verso é contar a primeira vírgula como se constituísse uma sílaba.

Primeiramente, é preciso esclarecer que contar a vírgula — ou qualquer outro sinal de pontuação — como uma sílaba é pouco comum. Entre os mais de 100.000 versos processados nos experimentos desta tese, ape-

nas alguns poucos do *corpus* SIL, considerados inescandíveis, poderiam ser escandidos caso a pontuação contasse como sílaba. No seguinte verso, por exemplo, um dos travessões poderia contar como uma sílaba para que se obtivessem as 9 sílabas exigidas pelo metro do poema:

SIL:5.261.3.1 — Amor — só às almas ensina

Tal hipótese, contudo, é problemática, pois trata-se de um *corpus* com uma grande quantidade de versos inescandíveis (43 ou 3,7%), cuja ampla maioria não pode ser explicada desta mesma forma. Se há tantos outros versos que são inescandíveis provavelmente devido a erros de edição, é possível que também no verso acima haja algum erro. Ademais, contar o travessão como uma sílaba não resolve de forma definitiva a escansão do verso: todos os outros 19 versos do poema se encaixam no esquema rítmico 3-6-9, mas, se contarmos o segundo travessão como uma sílaba, o esquema rítmico não é compatível. Por outro lado, contar o primeiro travessão como pausa — o que faria o verso se encaixar no esquema rítmico — seria no mínimo curioso, pois se trata do início do primeiro verso de uma estrofe, e portanto já existe uma pausa antes mesmo de o verso iniciar.

Em segundo lugar, a pontuação realmente pode, como no caso do verso de Gonçalves Dias, encontrar-se numa posição em que se faz necessária a manutenção do hiato. Disso, contudo, não se pode concluir que toda pontuação ou mesmo que a maior parte da pontuação deva impedir a sinalefa. Não existe, de fato, nenhum sinal de pontuação que sempre impeça a sinalefa. A vírgula não tem esse poder:

1 2 3 4 5 6 7 8 9 10 11  
MAT:1.2.1.2 In-fes-te-to-da a-te-rra, in-va-da os-ma-res,

Tampouco o ponto e vírgula, mesmo acompanhado do travessão, consegue evitar a sinalefa:

1 2 3 4 5 6 7 8 9 10 11  
MAG.4.16.3 A-Ra-zão,és-tu mes-mo; — a-li-ber-da-de,

Nem o ponto de interrogação, nem o de exclamação, nem os parênteses, nem as reticências:

1 2 3 4 5 6 7 8 9 10 11 12  
COS.13.1.1 Ni-se? Ni-se? on-de es-tá-s? A-on-de es-pe-ra  
PIN.2.3.2 Va-lei-me! A-qui, ó men-te, que guar-das-te  
CAM.4.13.8 Ne-ga-rão (co-mo Pe-dro) o Deus que têm.  
DIA.2.26.8 A-rran-ca e ro-ça a fe-ra!... a fe-ra a-tô-ni-ta,

Provar, entretanto, que nenhuma pontuação é capaz de impedir a sinalefa não diz nada a respeito da frequência com que a pontuação coincide com um hiato que normalmente seria desfeito por uma sinalefa. Por outro lado, se a premissa do sistema estivesse profundamente

equivocada, a taxa de acerto obtida nos experimentos não seria tão alta. Ademais, é importante notar que remover a pontuação não implica, de forma alguma, a sinalefa. Aquele verso de Gonçalves Dias discutido por Manuel Bandeira, por exemplo, é escandido assim pelo sistema:

1    2    3    4    5    6    7    8    9    10   11  
DIA.3.22.3 Gri-|ta;|e os|seus,|me-|dro-|sos,|re-|cei-|an-|do

### 5.3. Detecção da vogal tônica

Esta seção descreve como o sistema determina a posição da vogal tônica (ou sua ausência) de palavras escritas em português. Para fazê-lo, o sistema utiliza regras que comparam as palavras com certos padrões e, com base na comparação, estabelece a posição do acento. Estas regras foram estabelecidas pelo próprio autor da tese tratando os textos escritos segundo a ortografia vigente como algo a ser explorado: em vez de buscar possíveis manuais que ensinem como saber onde está a vogal tônica de uma palavra, optou-se por observar as palavras do português e deduzir como o sistema ortográfico funciona. Documentos como o acordo ortográfico (67), embora úteis à tese, são escritos para demonstrar como se escrevem as palavras que já se sabe falar, não como devem ser lidas as palavras que já se sabe escrever.

**Ortografia do português.** A ortografia do português está repleta de ambigüidades, mas uma virtude indiscutível que ela possui é possibilitar ao leitor encontrar de forma determinística a vogal tônica da amplíssima maioria das palavras. Saber onde está o acento de uma palavra é essencial à escansão, já que os ictos dos versos procedem, de modo geral, dos acentos lexicais. Deve-se ressaltar que nem todas as ortografias são tão benévolas com os leitores, humanos ou robóticos. O inglês e o russo exigem que o leitor saiba com antecedência qual sílaba acentuar em cada palavra; o italiano é mais regular, mas as proparoxítonas não são marcadas (*subito, ultimo*); em tcheco e francês o acento é previsível (na primeira e na última sílaba, respectivamente). Exemplos de línguas cujo acento é refletido na ortografia são o espanhol, com regras similares à do português, e o grego (moderno e antigo), que usa diacríticos em toda palavra acentuada.

A ortografia contemporânea do português parte do princípio de que as paroxítonas são mais comuns e portanto não precisam ser marcadas com diacríticos. As oxítonas e as proparoxítonas recebem um agudo ou circunflexo, que, além da posição do acento, distingue também o caráter

aberto ou fechado de certas vogais, como em *bisavó* e *bisavô*. As palavras sem agudo ou circunflexo podem ser oxítonas ou paroxítonas, conforme certos padrões ortográficos. As regras do sistema utilizam os diacríticos e os arranjos dos grafemas para encontrar o acento.

**Regras.** O sistema dispõe de quatro regras, cada uma com suas sub-regras, que localizam o grafema vocálico tônico de uma palavra. As regras são aplicadas em sequência e o sistema considera como final o resultado da primeira que apontar o acento. As regras são:

- Regra E: as exceções;
- Regra D: palavras com diacríticos determinantes de acento;
- Regra F: palavras cujos grafemas consonânticos finais restringem a posição do acento;
- Regra C: palavras que somente uma análise mais complexa dos grafemas vocálicos pode revelar a posição do acento.

A Tabela 5.1 enumera as regras e sub-regras. Cada sub-regra é acompanhada de informações coletadas nos *corpora* LIT e HOU: um exemplo de aplicação, o número de palavras em que ela foi utilizada e a frequência relativa correspondente.

O Aoidos começa o processamento removendo os diacríticos trema e grave de palavras como *tranqüilo* e *àquele*. Nenhum destes dois diacríticos influencia a posição do acento; o grave sequer altera a pronúncia de qualquer palavra — ao menos no Brasil. Apesar de o trema não ser mais usado após a adoção do novo acordo ortográfico, o Aoidos propõe-se a processar textos escritos também no padrão anterior, e portanto não rejeita palavras com este diacrítico.

Em seguida o sistema passa à Regra E: a lista de exceções. O objetivo desta lista não é reunir casos difíceis ou corrigir de maneira forçada os erros do sistema; seu objetivo é coletar os casos que a ortografia em si não é capaz de tratar. São quatro as sub-regras de exceção que o sistema considera. Na primeira, E1, há apenas a palavra *sob*: é a única palavra de uso geral que termine em *b* empregada em português com qualquer frequência significativa. A ortografia não dispõe de meios para explicitar a ausência de acento na palavra; o sistema a considera átona por exceção. A frequência absoluta de aplicação desta sub-regra no *corpus* HOU, conforme a Tabela 5.1, é dois, pois a sub-regra foi aplicada na palavra hifenizada *sob-roda*.

Tabela 5.1.: Regras para a localização do grafema vocálico tônico.

Regra	LIT			HOU		
	Exemplo	Absoluta	Relativa	Exemplo	Absoluta	Relativa
E1	<i>sob</i>	911	0,03%	<i>sob</i>	2	0,00%
E2	<i>ao</i>	96.479	3,50%	<i>com</i>	15	0,01%
E3	<i>semi</i>	90	0,00%	<i>hiper</i>	182	0,14%
E4	<i>ah</i>	3.858	0,14%	<i>vapt</i>	8	0,01%
		101.338	3,67%		207	0,16%
D1	<i>infância</i>	204.710	7,42%	<i>doméstica</i>	27.080	20,80%
D2	<i>manhã</i>	3.029	0,11%	<i>cardigã</i>	408	0,31%
		207.739	7,53%		27.448	21,11%
F1	<i>assaz</i>	165.423	6,00%	<i>abdominal</i>	20.566	15,80%
F2	-	2	0,00%	<i>trips</i>	1	0,00%
F3	<i>festins</i>	2.948	0,11%	<i>jardins</i>	12	0,01%
F4	<i>algum</i>	49.566	1,80%	<i>mirim</i>	983	0,76%
		217.939	7,90%		21.562	16,56%
C11	<i>colibri</i>	19.907	0,72%	<i>caju</i>	2.010	1,54%
C12	<i>bens</i>	26.866	0,97%	<i>quem</i>	168	0,13%
C13	<i>que</i>	767.151	27,80%	<i>do</i>	2.765	2,12%
C14	<i>alegre</i>	978.652	35,47%	<i>moscas</i>	52.752	40,52%
C15	<i>radiante</i>	24.262	0,88%	<i>raíinha</i>	1.627	1,25%
C16	<i>Bocaiuva</i>	2	0,00%	<i>gandaiice</i>	40	0,03%
C17	<i>noite</i>	70.241	2,55%	<i>ouro</i>	5.626	4,32%
C18	<i>meteoro</i>	22.964	0,83%	<i>italiana</i>	3.003	2,31%
		1.910.045	69,23%		67.991	52,23%
C21	-	4	0,00%	<i>tapii</i>	7	0,01%
C22	<i>adeus</i>	299.581	10,86%	<i>reação</i>	11.718	9,00%
		299.585	10,86%		11.725	9,01%
C31	<i>ancião</i>	6.999	0,25%	<i>fi<u>au</u></i>	327	0,25%
C32	<i>ensaio</i>	14.419	0,52%	<i>aleluia</i>	880	0,68%
C33	-	7	0,00%	<i>mal<u>ai</u></i>	3	0,00%
C34	<i>ca<u>iu</u></i>	996	0,04%	-	0	0,00%
		22.421	0,81%		1.210	0,93%
		2.759.067	100,00%		130.183	100,00%



A segunda sub-regra de exceções, E2, contém sete palavras átonas: *por, porque, em, ao, aos, ou e com*. A cada uma destas palavras podemos encontrar outra, bastante similar, que a ortografia considera tônica: *cor, torque, cem, tao, caos, vou e tom*. O sistema considera essas palavras excepcionais, já que não é, por exemplo o *p* que faz o vocábulo *por* átono e o *c* que faz *cor* tônico. Cumpre observar que este *porque* é o “de reposta”; o “de pergunta” seria escrito *por que* ou *por quê* e o substantivo, *porquê*. A frequência absoluta desta regra para o HOU também é maior do que o número de exceções por causa de palavras hifenizadas como *maria-com-a-vovó* e *amor-em-penca*.

O terceiro grupo de exceções, tratados pela sub-regra E3, são aqueles prefixos passíveis de serem seguidos por hífen em palavras compostas. Eles foram considerados exceções porque fogem à regra geral de que o hífen é usado para separar componentes cuja tonicidade ou atonicidade pode ser determinada de forma independente. Os prefixos escritos da mesma forma que palavras independentes não constam desta lista. Esta lista até o presente momento contém 21 prefixos, a saber: *anti, circum, hiper, intra, super, arquí, inter, mini, multi, pluri, proto, eletro, bio, geo, macro, micro, neo, pan, pseudo, retro e semi*.

A quarta e última sub-regra de exceções, E4, contém interjeições e palavras de origem onomatopeicas, que frequentemente não seguem os mesmos princípios ortográficos das demais palavras. Estão presentes na lista quatro palavras: *oh, ah, vapt e vupt*.

A Regra E tem uma frequência relativa de 0,16% no *corpus* HOU e de 3,67% no LIT, o que põe em evidência a diferente natureza dos dois *corpora*: no dicionário são poucas as palavras excepcionais, mas elas tendem a ser usadas na prática com uma boa frequência. A única sub-regra anômala é a E3: ela trata os prefixos seguidos de hífen, que tendem a ser palavras técnicas e portanto pouco presentes na prosa e na poesia.

Com as exceções devidamente tratadas, o Aoidos faz três tentativas sucessivas de encontrar a vogal tônica: através da presença de um diacrítico (Regra D), por meio da consoante final da palavra (Regra F) e com uma análise dos grafemas vocálicos na parte final da palavra (Regra C).

A Regra D1 procura por um agudo ou circunflexo, diacríticos estes que mantêm uma correspondência unívoca com o acento: o grafema vocálico ao qual eles se sobrepõem sempre leva o acento da palavra. São exemplos *ábaco, bocó, relâmpago, patê*. O til não é levado em conta na Regra D1, pois tem como função primordial indicar a nasalidade do grafema vocálico correspondente: *pão, põe*. Ele pode, inclusive, aparecer ao lado do agudo

e do circunflexo, sem influenciar no acento: *órgão*, *bênção*. A única situação em que o til determina o acento, que é tratada pela Regra D2, surge em palavras terminadas em *ã* ou *ãs* que não contenham agudo ou circunflexo: *maçã*, *cortesã*. Nas demais palavras sem agudo ou circunflexo, o til apenas coincide em geral com o acento: *mão*, *cãibra*. No entanto, existem numerosas palavras, sobretudo derivadas, em que o acento encontra-se em posição diversa da do til: *cãozinho*, *mamãezinha*. Uma possível exceção não tratada pelo sistema é a palavra *oçaim*.

A Regra F, assumindo que as regras anteriores trataram as palavras que contêm diacrítico determinante de acento, examina as consoantes finais, que podem atrair o acento para fim da palavra. A primeira sub-regra, F1, procura por um grafema consonântico dentre *l*, *n*, *r*, *x* e *z* no fim da palavra, precedido de uma dos grafemas vocálicos *a*, *e*, *i*, *o* ou *u*, e atribui o acento ao grafema vocálico. Exemplos de palavras tratadas por esta sub-regra que terminem em *l*, *r* e *z* são muitos, como *coronel*, *coroar*, *talvez*. As palavras terminadas em *x* são raras, mas podem ser encontradas, como *petromax*, *duplex*, *remix*, *solovox*, *radiolux*. Palavras relevantes terminadas em *n*, contudo, não foram encontradas, apenas contraexemplos que, necessitando de acento, comprovam a regra: *brâman*, *pólen*, *písin*, *bóson*. Nenhum contraexemplo terminado em *un* foi encontrado.

A Regra F2 busca pela terminação *ps*. A existência desta regra é inferida através de alguns poucos contraexemplos: *bíceps*, *tricerátops*. Partindo do princípio de que tais palavras não levariam diacrítico à toa, a Regra F2 posiciona o acento antes da terminação *ps*. Contudo, a única palavra que foi processada pela regra nos *corpora* LIT e HOU, *trips*, poderia ter a posição de seu acento determinada de forma mais simples. Esta regra, portanto, foi incluída mais por questões de completude do que por outras razões. A Tabela 5.1 informa que 2 palavras foram encontradas no LIT. São duas instâncias da palavra inglesa *perhaps*, que não servem de exemplo para o português.

As palavras terminadas em *ns*, quando precedidas de *i*, *o* ou *u*, são tratadas pela Regra F3. Exemplos de tais palavras, que recebem o acento no último grafema vocálico, são *patins*, *bombons* e *ticuns*. Sabe-se que a terminação *ens* não atrai o acento através de inúmeros contraexemplos, como *convéns* e *itens*. Não foram encontradas evidências conclusivas para a terminação *ans*, mas, com base em palavras questionáveis como *dólmans* e *Nova Orleans*, do *corpus* LIT, a terminação parece de fato atrair o acento.

A sub-regra F4, similarmente, acolhe as palavras terminadas em *im*, *om* e *um*, atribuindo-lhes o acento no último grafema vocálico. Exemplos

Tabela 5.2.: Exemplo de seções produzidas para a regra C.

Palavra	–	V1	C	V2	MN	–
<i>pr<u>a</u>tos</i>	pr	a	t	o		s
<i>im<u>a</u>gem</i>	im	a	g	e	m	
<i>cheq<u>e</u></i>	ch	e	qu	e		
<i>aind<u>a</u></i>	a	i	nd	a		
<i>ca<u>i</u>xa</i>	c	ai	x	a		
<i>nevo<u>e</u>iro</i>	nev	oei	r	o		
<i>caie<u>r</u>a</i>	c	aiei	r	a		
<i>pi<u>a</u>ui<u>e</u>nse</i>	p	iaue	ns	e		
<i>caix<u>a</u>o</i>	c	ai	x	ão		
<i>torne<u>i</u>o</i>	t	o	rn	eio		
<i>urugu<u>a</u>io</i>	ur	u	g	uaio		
<i>ur<u>s</u>o</i>		u	rs	o		
<i>t<u>i</u>o</i>			t	io		
<i>a<u>i</u>a</i>				aia		

de palavras às quais se aplica a regra são *amorim*, *garçom* e *batecum*. Contraexemplos relevantes são *amam* e *correm*.

Tratados os casos excepcionais, com diacríticos determinantes de acento e com consoantes atrativas de acento, resta uma série de casos mais complexos. Para processá-los, o Aoidos remove os possíveis tis sobreviventes e particiona a palavra em seis seções, como mostra a Tabela 5.2: uma sequência inicial de grafemas quaisquer, que é ignorada; a seção V1, composta de uma ou mais grafemas vocálicos; a seção C, correspondente a um ou mais grafemas consonânticos; a seção V2, semelhantemente feita de um ou mais grafemas vocálicos; a seção MN, uma sequência vazia ou então o grafema *m* ou *n*; um sufixo, que pode ser nulo ou o grafema *s*, mas que, de qualquer maneira, é ignorado. Este particionamento trata o grafema *u* das sequências *qu* e *gu* como grafemas consonânticos, pois ou ele forma dígrafo, como em *arquear* e *erguer*, ou eles correspondem a semi-vogais silabicamente presas à consoante precedente, como em *cinquenta* e *aguentar*.

A partir desse particionamento, averigua-se a quantidade de grafemas na seção V2. Se houver dois ou mais grafemas vocálicos, o acento recairá sobre um desses grafemas; se houver apenas um, o acento pode encontrar-se na seção V2 ou na seção V1. As sub-regras cujos nomes começam com C1 tratam do caso quando V2 possui um grafema; C2, quando possui dois; C3, quando possui três ou mais.

Quando a seção V2 contém apenas um grafema vocálico, o Aoidos verifica, através da Regra C11, se esse grafema é *u* ou *i*; caso seja, o acento é atribuído a ele, como em *uirapuru* e *caqui*. Se a seção MN não estiver vazia, a Regra C12 afirma que se trata de uma palavra monossilábica tônica, como *vem*, *cem*; palavras polissilábicas com terminações semelhantes, como *convém* e *partem*, não chegariam até aqui, pois Regras como a D1 ou F4 já teriam cuidado delas. A Regra C13 diz que, se a seção V1 estiver vazia, o sistema está diante de um monossílabo átono, como *te*, *pra*, *do*. Tais palavras, para serem tônicas, necessitariam de um diacrítico. As palavras com um grafema na seção V2 que restam todas recebem o acento em algum grafema vocálico da seção V1; resta determinar qual. Se houver apenas um grafema em V1, ele é o acentuado pela Regra C14, como *vista*, *velhiçe*. Se a seção C contiver dois ou mais grafemas consonânticos e o último deles for *l*, *m*, *n* ou *r*, então esse grupo de grafemas atrai o acento para o grafema vocálico que o procede: é a regra C15. São exemplos as palavras *hiulco*, *Coimbra*, *ainda*, *demiurgo*. Sabe-se que não é o primeiro grafema consonântico de tais grupos que atrai o acento: *braile*, *açaimo*, *plaino*, e *terneiro*. Também não são quaisquer grupos de dois grafemas consonânticos, como atestam *exausto*, *saibro*, *laicra*. A Regra C16 especifica que, se houver três ou mais grafemas vocálicos em V1 e os dois últimos forem *ui*, *iu*, *ii* ou *uu*, o acento se encontra no último grafema da seção V1; exemplos são *baiuca*, *cauila*, *gandaiiçe* e *auuva*. Quando V1 terminar em *u* ou *i* (porém não em *uu* ou *ii*), o acento recua para o grafema vocálico precedente; a Regra C17 cuida, portanto, de palavras como *abaxo* e *circuito*. As palavras que porventura não tiverem sucumbido a nenhuma das Regras C11 a C17, terão o acento atribuído pela Regra C18, que o coloca no último grafema de V1: *coala*, *bucuuva*, *xiita*.

Se houver dois grafemas vocálicos na seção V2, o acento cai em um deles. De acordo com a Regra C21, se esses dois grafemas forem *ii* ou *uu*, então o acento está no último: entre os raros exemplos estão *tapii* e *nhambuu*. As quatro instâncias no *corpus* LIT a que se refere a Tabela 5.1 são as palavras latinas *filii* e *calcii* (esta duas vezes) e aquilo que presumivelmente se trata do plural do nome da letra *u*, escrito *uu*; nenhum desses casos serve de real exemplo da regra. Em todos os outros casos, a Regra C22 diz que o acento está localizado no primeiro grafema vocálico, como em *tentei*, *andai*, *apatia*, *acentua*, *pessoa*.

Quando houver três ou mais grafemas vocálicos em V2, entram em ação as Regras C31 a C34. A primeira delas, C31, afirma que, se o grafema vocálico médio for *a*, *e* ou *o*, então é justamente nele que recai o acento,

como em *lea*, *lua*, *mia*, *tatuei*, *tateo*. Se, contudo, o grafema médio for *i* ou *u* e o grafema final for *a*, *e* ou *o*, pela Regra C32 o acento está no primeiro grafema da seção V2, como em *caia*, *jaminaua*, *conluio*, *baniua*. A Regra C33 examina os dois últimos grafemas vocálicos da seção V2: se forem *ui*, o acento similarmente encontra-se no primeiro grafema da referida seção, como em *malaui*, *matanaui* e *saaraui*. As instâncias de aplicação desta regra no *corpus* LIT são, conforme a Tabela 5.1, sete; porém uma delas é o nome próprio francês *Louis* e as outras seis são a palavra *pauis*, erroneamente acentuadas pelo sistema. A última sub-regra, C34, trata os casos restantes, que terminam em *iu* e recebem o acento no grafema médio da seção V2; tais palavras são as formas verbais da 3ª pessoa do singular do pretérito perfeito de verbos cujo infinitivo termina em *air* ou *uir*: *caiu*, *evoluiu*. O fato de que apenas tais formas verbais são processadas pela Regra C34 explica por que precisamente nenhum caso foi encontrado no *corpus* HOU: o dicionário não contém formas flexionadas.

**Limitações.** Existem palavras que conhecidamente são tratadas de forma errônea pelas regras do sistema. O autor optou por não considerá-las como exceções porque não pertencem a classes fechadas. As formas problemáticas conhecidas são as seguintes:

- Certas flexões de verbos que expõem a sequência *qu* ou *gu* às desinências. Exemplos são a cacofônica 3ª pessoa do singular do subjuntivo presente do verbo *averiguar*, *averigue*, que o sistema acentua *averigue*, e a bizarra 3ª pessoa do singular do presente indicativa do verbo *delinquir*, *delinqui*, que o sistema acentua *delinqui*.
- O plural de palavras terminadas em grafema vocálico mais *ul*. O sistema faz delas paroxítonas, quando elas são, na verdade, oxítonas. As palavras conhecidas que se encaixam nessa descrição são *paul*, *baul*, *miul* e *peul*, cujos plurais corretos são *pauis*, *bauis*, *miuis* e *peuis*, e não *\*pauis*, *\*bauis*, *\*miuis* e *\*peuis*, como quer o sistema. O problema aqui é que tais palavras estão em conflito com outras como *malaui* e *matanaui*. A ortografia atual não permite distinguir um caso de outro.
- As palavras *uaiua* e *eoipo*, que o sistema acentua *uaiua* e *eoipo*.

Além dos erros conhecidos, existem palavras que, embora o sistema as considere tônicas, elas nem sempre o são. Exemplos notáveis são

as formas *um* e *uma*, que, na função de artigo, dificilmente são acentuadas, como em *um carro*, *uma casa*; todavia, quando elas exercem outras funções, como a de numeral em *um ou dois* e *uma ou duas*, podem ser acentuadas. Tampouco são levados em consideração os acentos secundários, que existem sobretudo em advérbios derivados de adjetivos, como em *sing[e]lamente*, *pac[i]ficamente*.

Há também as palavras que, embora escritas corretamente em português, contêm grafemas não utilizados nas palavras nativas, e por esse motivo o sistema covardemente se recusa a processá-los, produzindo um erro. Palavras assim são *darwinismo*, *cyrillácea*, *stackhousiácea*, *baddeleyita*. Não seria, com efeito, difícil alterar o sistema para que palavras assim se encaixassem nas regras descritas nesta seção. Tais palavras, contudo, são parte de um problema maior, o das palavras e nomes estrangeiros, que não pode ser tão facilmente resolvido. Nos seguintes versos, como o sistema pode saber que deve pronunciar *kinnors*, *Macbeths* e *Vinci*?

ANJ.27.8.2 Troam kinnors! Depois tudo é tranquilo...

ANJ.1.22.3 — Macbeths da patológica vigília,

ANJ.12.29.2 Que eu descobri, maior talvez que Vinci,

#### 5.4. Divisão silábica

O Aoidos, a partir do conhecimento da posição da vogal tônica de uma palavra, divide-a em sílabas por meio de duas etapas: uma que classifica cada grafema em vogal ou consoante, e outra que faz a divisão propriamente dita. A separação silábica final tem o objetivo de corresponder àquela dos dicionários: trata-se de uma pronúncia conservadora e pausada. As características mais fluidas da língua falada são incorporadas às palavras em etapas posteriores do Aoidos.

A divisão silábica parte do princípio de que as vogais são núcleos de sílabas e que, portanto, dividir um vocábulo em suas sílabas consiste em determinar quais de seus grafemas são vogais e agrupar as consoantes e semivogais com as vogais corretas. Por se tratar de uma divisão silábica que busca, neste estágio, ser ortográfica e não fonética, o acordo ortográfico de 1990 guiou a separação dos encontros consonantais (67).

**Vogais, consoantes e semivogais.** Esta etapa classifica cada grafema da palavra de entrada em vogal (V), consoante (C) ou semivogal (S). Por exemplo, para palavras como *maracatu*, *guardiães* e *abstração* são geradas as classificações CVCVCVCV, CSVCCVVC e VCCCCVCVS. A classifi-

cação é feita substituindo incrementalmente os caracteres da sequência de entrada com esses símbolos. A palavra de entrada é inicialmente convertida para minúsculas.

A classificação começa tratando de duas exceções: a forma *ao* e seu plural *aos*. Além de serem excepcionais no que se refere ao acento (são átonas), são excepcionais também porque o ditongo /aw/ é normalmente representado por *au*, como em *caule* e *mingau*. A estas duas formas são, portanto, atribuídas as classificações VS e VSC.

Com as exceções já processadas, todas as demais palavras passam pelo mesmo caminho. De forma geral, os grafemas vão sendo classificados em vogais, consoantes e semivogais até que restem apenas certos grupos ambíguos de *i* e *u*, que são classificados segundo uma heurística.

Os primeiros grafemas a serem tratados são aqueles das sequências que começam com *g* ou *q* e são seguidos de *u*, com ou sem trema. O sistema, por não comportar lista de exceções para palavras funcionais, simplesmente assume que o *u* é mudo nas sequências compostas de *a*) os grafemas *g* ou *q*, *b*) o grafema *u* sem trema, e *c*) os grafemas *e* ou *i*, com ou sem diacríticos. Assim, ele acerta em palavras como *queijo* e *mangueira*, mas erra em *frequência* e *língua*. Analisando-se todas as entradas do Houaiss (*corpus* HOU), observa-se que 5.583 delas (4,7%) contém tais sequências. A partir da informação de ortoépia dessas entradas, descobre-se que em apenas 5,9% delas a pronúncia do *u* é obrigatória e em 1,2% dos casos sua pronúncia é opcional. Portanto, estas sequências são relativamente raras e a pronúncia do *u* também. Os experimentos do próximo capítulo trazem dados mais concretos sobre o impacto desta decisão na taxa de acerto do sistema.

Seguindo essa decisão, o sistema substitui o *u* em tais sequências pelo grafema precedente, convertendo *queijo* e *mangueira* em *qqueijo* e *manggeira*. Essa conversão é utilizada somente nesta etapa de classificação dos grafemas e não é repassada aos estágios subsequentes. Um eventual *ü* é imediatamente classificado como semivogal; os grafemas consonantais *g* e *q*, duplicados ou não, não são classificados como consoantes agora porque regras posteriores dependem do grafema em si e não do símbolo C. As sequências *qu* ou *gu*, seguidas de *a* ou *u*, com ou sem diacríticos, tem seu elemento intermediário identificado como semivogal. Neste momento, palavras como *quero*, *sagüi* e *quatro* estão convertidas na representação interna assim: *qquero*, *sagüi* e *qSatro*.

Certos ditongos decrescentes podem já neste estágio ser reconhecidos. O grafema *ã*, quando seguido de *e*, *i* ou *o*, forma ditongo, como em

*mãe, cãibra e pão*. O mesmo vale para o grafema *õ*, que na ortografia vigente só é seguido pelo grafema *e*, formando também um ditongo, como em *pões*. Outro ditongo que pode ser identificado é a terminação *am* em certas desinências verbais, como em *cantam* e *saltam*. Se essas formas verbais utilizassem uma forma mais de acordo com a ortografia, deveriam ser escritas *\*cântão* e *\*sáltão*, a exemplo de *órgão* e *acórdão*. Todas essas sequências podem ser substituídas por VS na representação interna.

A divisão silábica das palavras chamadas “proparoxítonas aparentes” (67), como *área*, *cárie* e *vácuo*, pode ser feita de duas formas: uma seguindo a pronúncia prática das palavras (*á-rea*, *cá-rie*, *vá-cuo*) e outra mais tradicional (*á-re-a*, *cá-ri-e*, *vá-cu-o*). O Aoidos atualmente toma o primeiro caminho. Para fazê-lo, o sistema procura pelas seguintes sequências no fim da palavra, seguidas ou não de *s*: a) o grafema *a*, precedido de *e*, *i*, *o* ou *u*, como em *cesárea*, *tíbia*, *amêndoa* e *tábua*; b) o grafema *o*, precedido de *e*, *i* ou *u*, como em *gêmeo*, *pífio* e *mútuo*; c) o grafema *e*, precedido de *i* ou *u*, como em *planície* e *tênuê*. As terminações de palavras assim (excluído um possível *s* final) são marcadas como SV.

Determinados grafemas vocálicos podem ser classificados sem ambiguidade como vogais: são todos os grafemas *a*, *e* e *o*, com ou sem diacríticos. Possíveis casos de *e* e *o* semivocálicos já devem ter sido tratados anteriormente quando no fim de vocábulos. No interior, em geral tais grafemas não formam ditongos. Dentre os grafemas vocálicos, restam, portanto, somente *i* e *u*. Quando estes grafemas são as vogais tônicas, eles evidentemente não podem ser semivogais, como em *barril* e *Setúbal*. Uma última forma de determinar que *i* e *u* são vogais é observar se eles são seguidos de: a) *mp* ou *mb*; b) o grafema *n* mais um grafema consonântico qualquer, exceto *h*. Em tais circunstâncias, o *m* ou *n* tem a função de nasalização, o que forçosamente determina que são precedidos de vogal. Os grafemas *u* e *i* que não foram classificados até agora são substituídos por X na representação interna, para indicar que ainda não foram processados.

Todos os grafemas consonânticos são, finalmente, substituídos por C, e neste momento todos os grafemas da sequência original foram substituídos por C, V, S ou X. Resta apenas resolver as sequências de X.

Os grupos ambíguos de *u* e *i* são resolvidos com uma estratégia de alternância: a partir da definição do caráter vocálico ou semivocálico do primeiro grafema, os demais alternam entre um e outro. O primeiro grafema do grupo é classificado como uma semivogal se for precedido por um grafema que se sabe ser uma vogal; caso contrário, é classificado como vogal. Assim, em uma palavra como *m[i]galha*, o grupo ambíguo *i* é clas-



Tabela 5.3.: Grupos compostos dos grafemas *u* e *i*.

Tamanho	LIT		HOU	
	Total	%	Total	%
1	582,227	99,70%	80.718	99,28%
2	1,727	0,30%	563	0,69%
3	0	0,00%	18	0,02%
4	0	0,00%	1	0,00%

sificado como V; na palavra *ba[i]xo*, o grupo ambíguo *i* é classificado como S. Em *p[i]a[ui]ense*, o primeiro grupo ambíguo é classificado como V e o segundo como SV. Existem palavras com grupos maiores, mas a ampla maioria dos grupos é composto de apenas um grafema, como mostra a Tabela 5.3. Exemplos de palavras com grupos de três grafemas ambíguos são *b[uiu]çu*, *m[uiu]íra* e *a[iui]ú*; o único exemplo conhecido de palavra com quatro grafemas ambíguos em um grupo só é *t[uiui]ú*.

**Separação silábica.** A separação silábica final é obtida a partir dos grafemas que compõem a palavra e da classificação deles em vogais, semivogais e consoantes; a posição do acento não é utilizada. O sistema inicialmente faz uma separação provisória e então redistribui alguns grafemas consonantais.

A separação provisória é feita mantendo-se as consoantes sempre juntas à vogal que as segue — exceto no caso da última sílaba, em que as consoantes ficam presas à vogal precedente por não haver uma que as siga. As semivogais ficam presas à vogal mais próxima. Neste estágio, alguns exemplos de divisão provisória são *a-ba-ca-te*, *la-drão*, *quai-squer*, *a-bstracção* e *lo-rnhão*.

As consoantes que iniciam cada sílaba, exceto a primeira, são examinadas. O sistema busca certos grupos de consoantes que indiquem que as consoantes precedentes pertencem à sílaba anterior. Tais grupos são: a) um dos grafemas *p*, *b*, *c*, *g*, *t*, *d*, *f* ou *v*, seguido de *l* ou *r*, como em *o-[pr]ó-[br]io*, *A-l[fr]e-do* e *e-s[tr]a-ngu-lar*; b) os dígrafos terminados em *h*, como em *lo-r[nh]ão* e *mu-r[ch]o*. Se algum desses grupos for encontrado, todas as consoantes que o precedem são enviadas ao fim da sílaba anterior: *Al-[fr]e-do*, *es-[tr]a-ngu-lar*, *lor-[nh]ão*, *mur-[ch]o*, *abs-[tr]a-ção*. Caso nenhum desses grupos tenha sido encontrado, o sistema adota uma estratégia simples: grupos consonantais com mais de dois grafemas têm o

primeiro colocado no fim da sílaba anterior. Assim, a divisão final é obtida para palavras como *quai[s-qu]er* e *pá[s-s]a-ro*.

### 5.5. A máquina de regras

Até este ponto, as etapas do sistema trabalhavam com sequências de caracteres codificadas de acordo com a ortografia do português. A partir da próxima etapa, porém, o sistema passa a trabalhar com uma representação fonética. Para manipular a representação fonética, as etapas seguintes empregam a *máquina de regras*, cuja função é transformar uma estrutura de transcrição fonética através de uma especificação declarativa de alterações. A estrutura de transcrição fonética é aqui denominada *elocução* e a especificação formal, *regra*.

**A elocução.** Uma elocução representa uma possível maneira de pronunciar um dado verso ou palavra. Além de uma sequência de caracteres seq, que corresponde tipicamente — mas não sempre — à transcrição fonética, uma estrutura de elocução contém alguns outros campos que ajudam a aplicar as regras de forma eficiente e cautelosa. Estes campos estão associados à elocução de modo geral, a cada caractere ou a cada sílaba. Há um só campo global, *pts*: é a pontuação da elocução, uma medida relativa da sua naturalidade, que é manipulada pelas regras. Há dois campos associados aos caracteres:

- *ton*, que especifica se o caractere correspondente recebe um acento tônico ou não;
- *cvs*, que associa a cada caractere o qualificativo de consoante, vogal ou semivogal.

A cada sílaba são associados vários campos, a saber:

- *tam*, que é o tamanho da sílaba, contado em caracteres;
- *ini*, um booleano que determina se a sílaba em questão inicia uma nova palavra na sequência de caracteres;
- *fim*, um booleano que, similarmente, determina se a sílaba conclui uma palavra;
- *cong*, um booleano que marca a sílaba como congelada, isto é, uma sílaba que as regras não podem alterar;

- *barr*, um booleano que, quando verdadeiro, indica que após a sílaba existe uma barreira que as regras não podem cruzar. As regras podem alterar o que está antes ou depois da barreira, mas não ambos simultaneamente;
- *hier*, um inteiro que especifica que a sílaba só pode ser alterada por regras posicionadas acima de um determinado nível hierárquico.

As elocuições são inicializadas a partir das etapas anteriores. Para cada palavra é criada uma elocução, na qual *seq* são os grafemas da palavra, *ton* é preenchido de acordo com a posição do acento, em *cvs* são colocadas as classificações já obtidas em consoante, vogal e semivogal, *tam* é preenchido com o tamanho encontrado pela etapa de silabação, *ini* é verdadeiro somente para a primeira sílaba e *fim* somente para a última, e os demais campos booleanos são inicializados com falso e o inteiro com zero. A etapa de transcrição fonética, descrita na próxima seção, utiliza a máquina de regras para converter os grafemas em fonemas. As elocuições correspondentes a palavras são então concatenadas através de forma simples para formar a chamada *elocução básica* do verso, a partir da qual todas as demais são derivadas.

**As regras.** As regras contêm instruções de manipulação a serem aplicadas sobre as elocuições. A máquina de regras pode ou não conseguir aplicar a regra à elocução. Em caso de sucesso, uma nova elocução é criada com as alterações ditadas pela regra; em caso de falha, o resultado é nulo. As regras possuem vários parâmetros. Os parâmetros cujo nome começa com *b* são de busca, isto é, dão especificações daquilo que a máquina de regras deve procurar na elocução; aqueles cujo nome começa com *m* são de modificação e instruem quais alterações devem ser feitas na elocução. Os parâmetros que começam com *e* são usados somente em circunstâncias especiais.

Os parâmetros obrigatórios são apenas dois, enquanto vários outros são opcionais. Alguns dos parâmetros opcionais foram criados para estágios específicos do sistema, ainda que a máquina de regras em si não imponha nenhuma restrição com relação ao seu uso. De modo geral, existem os parâmetros de aplicação geral, aqueles que são utilizados pela etapa de transcrição fonética e aqueles que são usados pelo gerador de elocuições. No material que segue, os parâmetros específicos à transcrição fonética recebem um T sobrescrito, como em *pml<sup>T</sup>*; aqueles usados pelo gerador de elocuições recebem um G, como em *bfron*.

**Regras: parâmetros obrigatórios.** Os dois únicos parâmetros obrigatórios são:

- *ber*, uma expressão regular a ser buscada na sequência *seq* da elocução, que pode conter subexpressões definidas por parênteses e referências a elas. Uma regra de crase pode, por exemplo, especificar o padrão  $/([aɪv])\1/$ , para que as sequências de fonemas  $/aa/$ ,  $/ɪɪ/$  e  $/ʊʊ/$  sejam encontradas.
- *msub*, uma sequência de caracteres que toma o lugar do trecho encontrado pela expressão regular, podendo conter referências às subexpressões definidas por *ber*. No caso do exemplo anterior, *msub* poderia ser simplesmente  $/\1/$ , de maneira que  $/aa/$ ,  $/ɪɪ/$  e  $/ʊʊ/$  tornar-se-iam  $/a/$ ,  $/ɪ/$  e  $/ʊ/$ . Dois caracteres possuem interpretação especial em *msub*: *a*) um ponto indica uma quebra de sílaba, ou seja, a divisão da sílaba encontrada em duas, o que é necessário nas regras de diérese; e *b*) um apóstrofo indica a posição do fonema tônico, o que é preciso quando os fonemas são rearranjados por uma regra de tal modo que a posição do acento não pode ser prevista pela máquina de regras.

A máquina de regras calcula em quais sílabas se encontra o trecho encontrado pela expressão regular. Estas sílabas podem ter suas características examinadas pelos parâmetros de busca e alteradas pelos de modificação.

**Regras: parâmetros opcionais de busca.** Apenas com os parâmetros obrigatórios, a máquina de regras é extremamente limitada. Os parâmetros de busca restringem os casos em que as regras são aplicadas, permitindo-as serem mais específicas de um modo relevante aos problemas tratados pelo sistema. São os parâmetros opcionais de busca:

- *bqsil*, que determina a quantidade de sílabas que a expressão regular *ber* deve cobrir. Uma mesma expressão regular pode cobrir números diferentes de sílabas conforme o contexto. Uma regra de nasalização na transcrição fonética, por exemplo, pode buscar pela expressão regular  $/a[mn]/$ , que será encontrada tanto em *acampar* quanto em *amarelo*, mas no primeiro caso corresponderia a uma sílaba e no segundo, a duas.
- *bpar*, um inteiro que indica o índice da subexpressão em *ber* que contém o trecho sendo de fato alterado na elocução, separando-o do

que é meramente contexto. Caso o parâmetro não seja especificado pela regra, a máquina assume que o trecho inteiro encontrado pela expressão regular está sendo alterado. Uma regra de aférese, por exemplo, pode especificar uma expressão regular  $/(e)([sz][bpm])/$  e o valor 1 para *bpar*, indicando que o fonema a ser alterado é a vogal e não as consoantes que servem de contexto. Ao ser aplicada a uma palavra como  $/ez'mal.ti/$  para transformá-la em  $/'zmal.ti/$ , apenas o  $/e/$  está sendo alterado pela regra.

- *bcvs*, através do qual a regra especifica o qualificativo de cada caractere do trecho encontrado pela expressão regular: consoante, vogal ou semivogal. Uma regra pode, por exemplo, converter os grafemas *E* e *I* no fonema  $/j/$ , se eles forem semivogais; para tal, a regra pode buscar pela expressão  $/[EI]/$  e especificar que o caractere encontrado deve ser uma semivogal.
- *bfron*, que especifica se o trecho encontrado pela expressão regular *ber* deve coincidir com as fronteiras silábicas, à esquerda, à direita, ou ambos. Isto significa que o comportamento normal do sistema é não levar em consideração essas fronteiras. Por exemplo, uma regra que procure por  $/lar/$  pode coincidir completamente com as fronteiras silábicas no caso de uma palavra como *fa-[lar]*, somente à esquerda em  $[la-r]ei-ra$  e somente à direita em *a-co-p[lar]*.
- *bton*, que define o padrãoônico que as sílabas encontradas pela regra devem seguir. O padrão constitui-se de um arranjo de tantas posições quantas forem as sílabas encontradas, cada posição contendo um dos seguintes valores: *a)* indiferente; *c)* tônica; *b)* átona em qualquer contexto; *d)* átona pré-tônica; *e)* átona pós-tônica. O sistema utiliza os campos *ton*, *ini*, *fim* para averiguar se o padrão está em acordo com o trecho encontrado na elocução. Uma regra de sinalefa pode especificar, por exemplo, que as sílabas encontradas devem ser átona, tônica e átona, e que portanto as átonas podem se tornar semivogais.
- *blim*, através do qual a regra restringe sua aplicação a determinadas posições silábicas. De maneira similar a *bton*, o padrão é um arranjo que é sobreposto pelo sistema às sílabas encontradas pela expressão regular. Cada item do arranjo especifica *a)* se a sílaba em questão deve ser a inicial, não deve ser a inicial ou tanto faz; *b)* se a sílaba deve ser a final, não deve ser a final ou tanto faz. Uma regra de

ectlipse que deseje remover a trava nasal da preposição *com*, por exemplo, pode buscar pela expressão regular */kõ/* e utilizar *blim* para restringir a busca à sílabas que são simultaneamente o início e o fim de palavras, evitando portanto encontrar palavras como *comparar*.

**Regras: parâmetros opcionais de alteração.** Estes são os parâmetros que definem como a máquina de regras deve alterar a elocução, após todos os critérios de busca terem sido atendidos:

- *mpts*, um inteiro a ser adicionado à pontuação *pts* da elocução, caso a regra seja aplicada com sucesso. Caso *mpts* não seja especificado, a máquina de regras não altera o campo *pts*, o que é conveniente para as regras da transcrição fonética, em que o conceito de naturalidade não faz muito sentido. As regras utilizadas pelo gerador de elocuições, por outro lado, precisam definir este parâmetro com cuidado para que as elocuições de um mesmo verso possam ser ordenados segundo seu grau de naturalidade.
- *mreaaj*, que define um reajuste para o tamanho de cada uma das sílabas encontradas. Com base neste parâmetro, a máquina de regras altera o campo *tam* da elocução. Uma regra de transcrição fonética que queira, por exemplo, converter os grafemas *SS* no fonema */s/* deve diminuir o tamanho da sílaba à qual o primeiro *S* pertence.
- *mdel*, que especifica sílabas a serem excluídas entre as encontradas pela expressão regular. A exclusão de sílabas se refere exclusivamente aos campos que guardam informações sobre elas, como *tam*; a cadeia de caracteres *seq* não é alterada por este parâmetro. Uma regra de síncope, por exemplo, pode utilizar *mdel* para indicar que uma sílaba no interior do vocábulo não existe mais, como a transformação de */fe.li.si'da.di/* em */fi.si'da.di/*.
- *mton*, que, para cada sílaba encontrada, especifica que *a)* sua acentuação não deve ser alterada; *b)* ela deve passar a ser tônica; ou *c)* ela deve passar a ser átona. Uma regra cujo objetivo seja, por exemplo, tornar tônica uma palavra átona como *que* quando em fim de verso, pode usar *mton* para indicar que a sílaba deve ganhar acento.
- *mmap*, que especifica mapeamentos a serem aplicados aos caracteres, uma vez que *seq* já tenha sido alterada pela máquina de regras. O objetivo deste parâmetro é dar uma maior expressividade às regras.

Uma regra de sinérese, por exemplo, que queira reunir uma vogal tônica mais uma átona numa mesma sílaba pode especificar uma expressão regular /a[ɪʊ]/ e valores apropriados para os parâmetros *bton*; os fonemas /ɪ/ e /ʊ/, entretanto, não devem ser transformados numa mesma semivogal e por isso o parâmetro *msub* não pode especificar diretamente que /ɪ/ deve tornar-se /j/ e /ʊ/, /w/. Cada mapeamento do parâmetro *mmap*, portanto, especifica: *a*) a posição em *ber* em que se encontra o caractere de origem do mapeamento; *b*) um mapa associando caracteres a caracteres, como  $\text{ɪ} \rightarrow \text{j}$  e  $\text{ʊ} \rightarrow \text{w}$ ; *c*) a posição em *msub* onde o resultado do mapeamento deve ser colocado.

**Regras: parâmetros especiais.** Existem regras bastante particulares que o gerador de elocuições possui: *a*) duas que aplicam fenômenos obrigatórios na fonologia do português, a sonorização de /s/ diante de consoante sonora, como *as mães*, que deve ser pronunciado /az'mãjs/, e a redistribuição de certas consoantes quando diante de vogais, como *as asas*, que é pronunciado /a'za.zas/, e *b*) uma regra que bloqueia o interior das palavras, impedindo que futuras sinalefas interfiram com sílabas que só poderiam ter sido previamente alteradas por meio de sinéreses. Os dois parâmetros especiais trabalham justamente com este bloqueio:

- *emcong*, que congela todas as sílabas internas das palavras, independentemente do padrão *ber*. As sílabas internas, neste contexto, são aquelas que: *a*) não estão em contato com palavras adjacentes, ou *b*) estão no início da palavra, mas começam por consoante, ou *c*) estão em fim de palavra, mas terminam por consoante. Por exemplo, em *a homenagem bela*, as sílabas *me*, *na* e *be* são internas, mas *a*, *ho*, *em* e *la* não; em *dois pães de trigo*, as sílabas *dois*, *pães* e *tri* são internas, mas *de* e *go* não. Ao receber este parâmetro, portanto, a máquina de regras ativa o campo *cong* de todas as sílabas internas e o campo *barr* da primeira sílaba dos dissílabos.
- *ebcong*, que instrui a máquina de regras a ignorar os valores do campo *cong*. As duas regras relativas aos fenômenos obrigatórios discutidos há pouco são as únicas a utilizarem este parâmetro. O que sucede é que o congelamento das sílabas internas precisa ser realizado *antes* dos fenômenos obrigatórios, porque esses últimos alteram as fronteiras entre as sílabas; por outro lado, esses fenômenos precisam alterar certas sílabas internas, como *nas* em *nas estre-las*, que se torna /na.zes'tre.las/. Diferentemente das demais regras,

contudo, estas dos fenômenos obrigatórios fazem as alterações de forma bastante controlada e que na prática não prejudicam o congelamento de sílabas.

**Controle de aplicação das regras.** Ao aplicar uma regra a uma elocução, a máquina de regras recebe dois parâmetros adicionais, externos à regra: o índice da ocorrência a ser buscada, *bn*, e o nível hierárquico da regra, *bmhier*.

O parâmetro *bn* faz-se necessário porque uma única regra pode ser aplicável em mais de uma posição na mesma elocução. No seguinte verso, duas sinéreses muito parecidas, uma em *esfaqueado* e outra em *criança*, precisam ser feitas para que o decassílabo exigido pela métrica do poema seja atingido:

1 2 3 4 5 6 7 8 9 10 11  
ANJ.12.17.3 O es-|tô-|ma-|go es-|fa-|quea-|do|de u-|ma|crian-|ça

Já neste outro verso, é preciso fazer uma sinérese em *diadema* ou em *real*, mas não em ambos:

1 2 3 4 5 6 7 8 9 10 11  
MAG.22.15.12 Com o|dia-|de-|ma-re-|al-|de-|pon-|do os|lou-|ros;

Tanto o vocábulo *diadema* é atestado com diérese alhures quanto *real* com sinérese:

1 2 3 4 5 6 7 8 9 10 11 12  
SIL:5.137.4.7 s o | di- | a- | de- | ma | da | vir- | tu- | de en- | can- | ta  
ANJ.174.2.1 No | bal- | da- | qui- | no a or- | ques- | tra | real | se a- | pres- | ta |

Se a mesma regra de sinérese for a responsável por todos esses casos, ela não pode ser aplicada indistintamente a todas as ocorrências de *ber* na elocução. O parâmetro, *bn*, portanto, pode ser usado para restringir a aplicação da regra somente à *n*-ésima potencial aplicação da regra dentro da elocução. Este controle é exercido pelo gerador de elocuições, visto mais adiante.

O segundo parâmetro, *bmhier*, é um inteiro que permite ao gerador de elocuições organizar as regras de forma hierárquica. Quando este parâmetro é passado à máquina de regras, o campo *hier* é alterado nas posições correspondentes às sílabas alteradas, de forma que tentativas de alterar aquelas sílabas no futuro só obterão sucesso se a máquina de regras receber um parâmetro *bmhier* maior que o utilizado anteriormente. Deste modo interferências entre regras do mesmo nível hierárquico são minimizadas.



**O contexto e a modificação.** A regra deve especificar o que é contexto e o que está efetivamente sendo alterado através de uma subexpressão em *ber* e do inteiro *bpar*. Isto é necessário porque a máquina de regras mantém um histórico das alterações que são feitas, ligando as elocução derivadas às originais e os caracteres derivados aos originais. Deste modo, é possível rastrear a origem de um fonema desde a sua presença no resultado final à palavra escrita, passando pela aplicação de todas as regras que o modificaram. Sem este rastreamento, o resultado final da escansão estaria preso aos símbolos fonéticos, sem possibilidade de alinhá-los ao texto original do verso. Em outras palavras, sem o rastreamento o usuário receberia somente uma escansão assim:

1 2 3 4 5 6 7  
MAT:1.17.15.1 wa|'su|kar|'za|sja|ka|'bow

Enquanto, para a maior parte dos propósitos, é mais útil a seguinte representação, possível graças ao rastreamento:

1 2 3 4 5 6 7  
MAT:1.17.15.1 O a-|cú-|car|já|se a-|ca-|bou?

**Evolução das elocuições.** Alguns campos das elocuições, conforme regras de diferentes etapas são aplicadas, vão perdendo o sentido. O campo *cvs*, por exemplo, perde totalmente sua utilidade após o fim da transcrição fonética e, mais do que isso, seu valor deteriora progressivamente, pois não é mais atualizado. Os limites entre as palavras, definido por *ini* e *fim*, também vão se dissolvendo conforme as regras do gerador de elocuições mesclam vogais de diferentes palavras. Trata-se, de fato, de uma dissolução fonológica, não apenas computacional; tome-se, por exemplo, o seguinte verso:

CAM.6.85.1 Mas já a amorosa estrela cintilava

Em sua elocução básica, faz sentido que as fronteiras entre as palavras estejam claramente demarcadas. Contudo, a partir do momento que a vogal final de *já* funde-se ao artigo *a* e ao início do adjetivo *amorosa*, não se pode mais determinar com precisão onde termina *já*, onde começa *amorosa* e onde está o artigo *a*:

1 2 3 4 5 6 7 8 9 10 11  
CAM.6.85.1 maz|'za|mo|'rɔ|zes|'tre|la|sɪ|ti|'la|va

O gerador de elocuições, portanto, precisa garantir que as regras, a partir de um certo ponto, não utilizem mais um parâmetro como *blim*. Para fazê-lo, o gerador pode utilizar, além de mecanismos específicos seus, também o parâmetro independente de regra *bmhier*.

## 5.6. Transcrição fonética

A etapa de transcrição fonética utiliza a máquina de regras para produzir uma pronúncia para as palavras dos versos. As elocuições são inicializadas com os resultados das etapas anteriores: pré-processamento, localização do acento e silabação. O único processamento adicional antes da criação das elocuições é a conversão de todos os grafemas em maiúsculas, de maneira a separar claramente no processo o que são grafemas originais, em maiúsculas, e o que são símbolos do Alfabeto Fonético Internacional, em minúsculas. A descrição a seguir segue também esta convenção.

A transcrição fonética por meio de regras é possível graças à natureza da ortografia contemporânea do português, que é parcialmente fonética. O sistema contém 81 regras para realizar a transcrição, descritas abaixo.

**Regras.** As regras iniciais aproveitam alguns resultados da etapa de silabação: (1) *E* e *I* em posição de semivogal tornam-se /j/, (2) *U* e *O* convertem-se em /w/; e os grupos (3) *GU* e (4) *QU*, quando ambos os grafemas estão classificados como consoantes, podem ser convertidos em /g/ e /k/.

Em posição inicial de palavra, (5) o *H* é removido, (6) o *R* muda-se em /ʁ/ e (7) o *X* em /ʃ/. Em posição final de palavra, (8) o grafema *Z* é convertido em /s/ e (9) o *X* em /ks/. (10) A sequência *ÊM* em posição é bastante particular: ela marca, apesar do agudo sobre o *E*, uma nasalização, e portanto é transformada em /ẽ/.

Alguns grafemas, com e sem diacrítico, podem ser agora transcritos. (11) O *A*, sem diacrítico ou portando o grave, agudo ou circunflexo é convertido em /a/, ainda que em alguns casos ele será sobrescrito mais tarde. (12) O grafema *A* com til torna-se o fonema /ã/, (13) *E* com agudo muda-se em /ɛ/ e (14) *E* com circunflexo em /e/. (15) O grafema *I*, com ou sem agudo, é transformado em /i/. De maneira similar ao *E*, (16) O grafema *O* com agudo converte-se em /ɔ/ e (17) com circunflexo em /o/. (18) Com til, o grafema *O* é mudado em /õ/. Por fim, (19) o *U*, com ou sem agudo, é transcrito em /u/.

Alguns casos particulares de consoantes são então tratados. (20) O grafema *S*, quando posicionado entre dois sons vocálicos (um dos fonemas /aãeeijooūw/) ou grafemas vocálicos (os grafemas *E* e *O* que ainda não foram transcritos), sonoriza-se em /z/. (21) O dígrafo *SS* é convertido em /s/ e (22) o dígrafo *RR* em /ʁ/.

A seguir são convertidas uma série de consoantes simples: (23) *B*

em /b/, (24) *D* em /d/, (25) *F* em /f/, (26), *J* em /ʒ/, (27) *P* em /p/, (28) *T* em /t/, (29) *V* em /v/ e (30) *Ç* em /s/.

O grafema *H* neste estágio compõe alguns dígrafos, de maneira que (31) *CH* é convertido em /ʃ/, (32) *LH* em /k/, e (33) *NH* em /ɲ/. (34) Um *H* que por algum erro no texto original ainda exista é nesse ponto removido.

Mais alguns grafemas simples são convertidos: (35) *L* em /l/, (36) *R* em /r/, (37) *Z* em /z/, (38) *M* em /m/ e (39) *N* em /n/, apesar de que a conversão de *M* ainda pode ser alterada pelas regras subsequentes.

Os grafemas *G* e *Q* são então tratados. (40) *G* é convertido em /ʒ/, contanto que seja seguido de um dos sons /eeij/ ou do grafema *E*; (41) qualquer *G* restante é mudado em /g/. (42) O grafema *Q*, seguido da semivogal /w/, torna-se /k/, e (43) seguido de /u/ também é transformada em /k/.

(44) Os dígrafos *XC* e (45) *XS*, contanto que separados em sílabas distintas pelas etapas anteriores, são convertidos em /s/, ficando na segunda das duas sílabas. (46) Também torna-se /s/ o *X* em fim de sílaba diante de qualquer som consonantal. (47) Casos de *X* que restem são convertidos em /ʃ/.

(48) O encontro consonantal *SC* e (49) o grafema *C*, seguidos de um dos sons vocálicos /eeij/ ou do grafema vocálico *E*, são transformados em /s/. (50) Qualquer *C* restante é então mudado em /k/ e (51) qualquer *S* sobrevivente torna-se /s/.

São resolvidas então as vogais nasais restantes. (52) A sequência de fonemas /am/ em fim de palavra muda-se em /ãw/; os fonemas /m/ e /n/ em fim de sílaba nasalizam então o grafema ou som vocálico precedente e desaparecem: (53) /a/ converte-se em /ã/, (54) /e/ ou *E* em /ẽ/, (55) /i/ em /ĩ/, (56) /o/ ou *O* em /õ/ e (57) /u/ em /ũ/. (58) O fonema /a/, em final de sílaba tônica e antes de /m/ ou /n/, é nasalizado em /ã/, como em *cama*.

Neste estágio resta apenas resolver os grafemas *E* e *O*, o que é particularmente difícil, visto que em muitos casos a ortografia vigente não permite distinguir entre vogais abertas (/ɛ/ e /ɔ/) e fechadas (/e/ e /o/). Existem, entretanto, alguns casos em que o contexto limita a distribuição das abertas e fechadas. Inicialmente, quando em sílaba tônica em fim de palavra e seguidas de /l/, (60) *E* torna-se /ɛ/ e (61) *O* torna-se /ɔ/, como em *mel* e *sol*. Em palavras átonas e seguidas de semivogal, (62) *E* é convertido em /e/ e (63) *O* em /ɔ/, como em *eu* e *ou*. Nas demais circunstâncias em palavras átonas (64) *E* é mudado em /ɪ/ e (65) *O* em /ʊ/, como em *de* e *do*. Diante do som /ɲ/, os grafemas vocálicos ficam fechados: (66) *E* em /e/ e (67) *O* em /o/, como em *penha* e *fronha*. Em posição pré-tônica, (68) *E* é mudado em /e/ e (69) *O* em /o/, como em *serraria* e *bordão*. Em posição fi-

nal de sílaba e seguidos de /m/ ou /n/, fecham-se, (70) *E* em /e/ e (71) *O* em /o/, como em *leme* e *tomo*. Em posição tônica e seguidas de /j/, também fecham-se, (72) *E* em /e/ e (73) *O* em /o/, como em *feio* e *coisa* (mas *ideia* e *boia*). Similarmente, tônicas e seguidas de /w/, são fechadas, (74) *E* em /e/ e (75) *O* em /o/, como em *celeuma* e *matadouro*. Os casos sobreviventes de *E* e *O* são dispensados da seguinte maneira: em sílaba tônica, ficam abertos, (76) *E* em /e/ e (77) *O* em /o/; em pós-tônica, (78) *E* é convertido em /i/ e (79) *O* em /u/; em posição pré-tônica, (80) *E* é transformado em /e/ e (81) *O* em /o/.

**Discussão.** As duas principais dificuldades na transcrição fonética do português são o grafema consonantal *X* e os grafemas vocálicos *E* e *O*. Sabe-se que em início de palavra *X* é sempre /ʃ/, mesmo em palavras de origem grega como *xantocromia*, que contrasta com *melanoxanto* em que o *X* é /ks/. No fim de palavra, é sempre /ks/. No interior das palavras, quando desacompanhado de outros grafemas consonantais, pode ser /z/, como em *exame*, /s/ como em *próximo*, /ʃ/ como em *feixe* e /ks/ como em *fixo*. Fazer tais distinções vai muito além do escopo deste trabalho.

A distinção entre as vogais abertas e fechadas, no caso de *E* e *O*, também é desafiadora. Nenhuma heurística que se confine às fronteiras da palavra conseguirá resolver todos os casos: pares como *eu olho* e *meu olho*, *sede* (vontade de beber água) e *sede* (unidade principal) o comprovam. Seria necessário realizar a análise sintática do texto, o que também vai além do escopo do trabalho.

Para propósitos de escansão, contudo, a transcrição fonética do sistema mostrou-se bastante eficaz, como evidenciam os experimentos descritos no próximo capítulo. As distinções entre vogais abertas e fechadas, embora destaquem-se em erros como /'es.ti/ para *este* e /'tɔ.dø/ para *tudo*, tiveram muito pouco impacto na escansão em si. O mesmo pode-se dizer a respeito do *X*: é estranho ver transcrições como /e'ʃis.ti/ para *existe*, mas isto em geral não influencia a escansão. Um caso particular em que o grafema *X* causou um problema prático de escansão foi no verso seguinte, encontrado durante os experimentos:

FIG:5.5.2.3 Não como frágeis, flexíveis vimes,

1	2	3	4	5	6	7	8	9	10
Não	co	mo	frá	geis,	fle	xí	veis	vi	mes,
'nāw	'ko	'mu	'fra	ʒɪjs	fle	'ʃi	vɪjz	'vi	mɪs

Este verso está num soneto decassílabo e a maneira correta de escandi-lo requer a realização de uma epêntese na palavra *flexíveis*, pronunciando-a

/fle.kɾ'si.vejs/. Embora o sistema seja capaz de executar tais regras (de fato, ele possui várias deste tipo), neste caso ele nada pode fazer pois o encontro consonantal /ks/ não está registrado em sua transcrição fonética.

## 5.7. Gerador de elocuições

O objetivo do gerador de elocuições é produzir uma série de maneiras alternativas de pronunciar um determinado verso, com o intuito de que entre elas estará a métrica ritmicamente correta. Para produzir as diferentes elocuições, o gerador parte de uma elocução base e cria outras elocuições aplicando regras de modificação que seguem uma estrutura hierárquica. A árvore de busca é explorada completamente ou até que um limite máximo de elocuições seja atingido. As elocuições não estão todas em pé de igualdade: elas contêm uma pontuação que corresponde à sua naturalidade, de sorte que umas elocuições são julgadas mais naturais pelo sistema do que outras. As elocuições mais *naturais* não são necessariamente as mais *corretas*, porém a tarefa de escolher a escansão mais apropriada para o contexto em que o verso aparece cabe às etapas seguintes do Aoidos.

**As camadas.** As regras utilizadas pelo gerador de elocuições estão organizadas em 10 camadas, indexadas por valores de 1 a 10. O propósito desta disposição hierárquica é produzir alternativas numa ordem em que as características das elocuições necessárias à correta aplicação das regras estejam intactas. As regras de elisão, por exemplo, observam o fim das palavras, mas as de sinalefa podem facilmente borrar a fronteira entre as palavras; logo, as elisões devem ser aplicadas antes das sinalefas.

O gerador mantém o registro de qual foi a regra que produziu uma elocução e somente permite que regras que pertençam a uma camada com índice igual ou superior à regra anterior sejam aplicadas. Assim, se uma elocução surgiu como resultado da aplicação de uma regra da camada 3, o gerador somente considerará aplicar a esta elocução regras das camadas com nível igual ou superior a 3, enquanto as regras das camadas 1 e 2 não podem mais ser aplicadas à elocução.

Quando uma regra de uma determinada camada produz uma elocução, outras regras da mesma camada podem ser aplicadas à mesma elocução. As regras de uma mesma camada, entretanto, não podem modificar as mesmas sílabas: uma vez que uma regra alterar uma sílaba qualquer, a máquina de regras ajusta o campo hier correspondente à sílaba e permite que apenas regras de uma camada de índice superior alterem a sílaba. O

gerador de elocuições informa, portanto, à máquina de regras qual é o índice da camada à qual pertence a regra sendo aplicada. Este controle exercido sobre a aplicação de regras difere, por exemplo, daquele de Gervás (98), em que metaplasmos como sinalefas são aplicados recursivamente na mesma sílaba.

Uma das camadas, a de índice 7, é especial: ela não pode ser pulada. Uma elocução produzida, por exemplo, por uma regra da camada 3 pode ser modificada por outras das camadas 3, 4, 5, 6 e 7, mas não 8 e 9. A camada 7 possui apenas uma regra, cujo objetivo é congelar as sílabas internas, utilizando para tal o parâmetro *emcong* da máquina de regras. Ela separa a hierarquia de camadas em duas partes: as de índice inferior a 7 podem especificar critérios que consideram o interior das palavras, e aquelas acima de 7 podem apenas olhar para as sílabas entre diferentes palavras que mantenham contato através de encontros vocálicos.

A camada 8 também tem suas particularidades: ela possui somente duas regras que são, de certo modo, obrigatórias. A primeira delas é responsável por tornar sonoro o fonema /s/ quando em fim de palavra e diante de vogal ou consoante sonoras, como nos dois exemplos seguintes:

MAT:33.21.2.9 traduzis em mau francês,

COS.111.1.2 Da parte mais vizinha, e mais distante,

No primeiro exemplo, o /s/ final de *traduzis* é pronunciado como /z/ por estar em fim de palavra e diante de uma vogal, /ẽ/. No segundo, o /s/ final das duas ocorrências de *mais* torna-se /z/ por causa das consoantes iniciais das palavras seguintes, ambas sonoras: /v/ e /d/. A segunda regra da camada 8 trata de um fenômeno igualmente compulsório: certas consoantes em fim de palavra passam, foneticamente, a pertencer à primeira palavra da sílaba seguinte, quando esta última começa com uma vogal. É o que acontece nos dois exemplos seguintes:

GON.7.8.9 Não diz uma fineza às tais mocinhas,

CAM.8.90.7 Lhe andar armando, que pôr em ventura

No primeiro, o /z/ no fim de *diz* — que era um /s/ antes da regra de sonorização — passa à palavra seguinte, *uma*. No segundo, os dois fonemas /r/ no fim dos infinitivos dos verbos passam às palavras seguintes, que começam por vogal. Os quatro exemplos das duas regras ficam assim foneticamente:

	1	2	3	4	5	6	7	8	9	10	11
MAT:33.21.2.9	tra	du	'zi	zẽ	'maw	frã	'ses				
COS.111.1.2	da	'par	tĩ	'majz	vi	'zi	ɲĩ	'majz	dis	'tã	tĩ
GON.7.8.9	nãw	'di	zu	ma	fi	'ne	zas	'tajz	mo	'si	ɲas
CAM.8.90.7	ʃjã	'da	'rar	'mã	du	ki	'po	rẽ	vẽ	'tu	ra

As duas regras da camada 8 são fonologicamente obrigatórias, porém o sistema não as aplica compulsoriamente como o faz com a regra da camada 7. O motivo é que não é necessário fazê-lo: basta que a pontuação das duas regras incentive seu uso e as elocuições que passaram em algum momento por elas sempre serão preferidas, em quaisquer contextos, em detrimento das que não passaram.

É possível que as camadas utilizadas pelo gerador de elocuições reflitam, de forma mais ou menos direta, processos linguísticos do português. Por exemplo, Bisol (60) diz que a ressilabação acontece após a elisão; em *camisa usada*, argumenta a autora, primeiro o /a/ de /ka.mi.za/ cai e depois o /z/ se junta ao /u/ da próxima sílaba, resultando na pronúncia *ka.mi.zu.za.da*. No Aoidos, de fato, a camada 8, que reposiciona a consoante /z/ em fim de palavra, aparece após a camada 6, de elisão. Estudos futuros poderão formalizar melhor a estrutura em camadas utilizada pelo Aoidos com princípios linguísticos.

**Naturalidade e penalidade.** As elocuições possuem um campo, *pts*, que tenta capturar a naturalidade da pronúncia por ela especificada. Na elocução básica, produzida a partir da junção das palavras individuais de um verso, o valor de *pts* é zero. Este não é um valor *neutro* de naturalidade, é um valor *inicial*. Esta pontuação, com efeito, é um valor relativo, não absoluto. Comparar a pontuação de duas elocuições produzidas para o mesmo verso faz sentido, mas comparar pontuações de elocuições oriundas de versos diferentes não.

As regras possuem um parâmetro, *mpts*, que, quando elas são aplicadas com sucesso, modificam a pontuação das elocuições. O parâmetro *mpts* pode ser encarado como uma penalidade: quanto mais extrema e incomum for uma regra, maior será a sua penalidade. Por outro lado, as regras que tornam as elocuições mais naturais possuem uma pontuação negativa: é uma penalidade ao contrário, um incentivo.

Uma consequência da abordagem de pontuação adotada pelo sistema é que a pontuação de versos maiores é frequentemente menor do que a de versos menores. O motivo é que certos fenômenos, como algumas sinalefas, tornam as elocuições mais naturais, e versos mais longos têm mais oportunidades de realizar tais sinalefas.

A melhor maneira encontrada na prática para comparar a pontuação de versos diferentes, sejam eles do mesmo tamanho ou não, foi tomar a menor pontuação entre todas as elocuições encontradas para um determinado verso como base para as comparações: em vez de utilizar o campo *pts*

de uma elocução, toma-se a menor pontuação encontrada,  $pts_m$  e calcula-se a diferença  $pts - pts_m$ .

**A busca.** A elocução inicial não é produto de nenhuma regra, mas sim da concatenação de elocuições menores, correspondentes às palavras do verso. O gerador de elocuições, portanto, atribui artificialmente a ela o índice de camada 1, de forma que todas as regras, até o nível 7, estão disponíveis à elocução básica.

A partir da elocução básica outras são derivadas através de um algoritmo de busca. O algoritmo deriva uma elocução de outras através da aplicação de regras, sempre respeitando a hierarquia de camadas. O conjunto de elocuições encontradas pelo algoritmo, portanto, naturalmente forma uma árvore em que os nós são elocuições e as arestas são as regras que derivam uma elocução de outra. Ciclos não existem na árvore porque o algoritmo, ao gerar uma elocução vista anteriormente, simplesmente a descarta. O algoritmo de busca mantém, deste modo, uma lista das elocuições presentes na árvore.

Para aumentar a árvore com novas elocuições, o algoritmo de busca escolhe um nó entre os existentes e tenta aplicar as regras disponíveis à elocução. Não é feita nenhuma escolha entre as regras da camada em questão: elas são tentadas uma a uma na ordem em que foram declaradas. Quando uma regra é aplicada com sucesso, a elocução resultante é adicionada à árvore — contanto que ele já não haja uma elocução igual, caso em que a busca no mesmo nó continua; duas elocuições são consideradas iguais pelo algoritmo se todos os seus atributos são iguais. Se nenhuma regra pode gerar uma nova e nunca antes vista elocução, o algoritmo marca o nó como completamente explorado e parte para outro. Um nó que receba essa marcação será para sempre um nó folha da árvore.

O algoritmo escolhe o nó a ser explorado a partir de todos os presentes na árvore, independentemente de sua localização. Os nós marcados como completamente explorados não são considerados na escolha. Para cada nó  $i$  cujas possibilidades ainda não foram esauridas o algoritmo calcula o índice

$$m_i = rs_i + pts,$$

em que  $s_i$  é o número de sílabas poéticas de  $i$  e  $r$  é um parâmetro de escala cujo valor é 10. O nó com menor valor  $m$ , sem critério de desempate, é então explorado pelo algoritmo. O número  $s$  de sílabas poéticas é utilizado, junto com o fator  $r$ , para dar preferência às elocuições com menos sílabas.



A motivação é a observação de que, em geral, é preciso reduzir o número de sílabas da elocução básica para obter a escansão correta.

Apesar de não escolher os nós de forma aleatória, o algoritmo faz uma busca exaustiva, só a interrompendo prematuramente caso o limite máximo de nós na árvore seja atingido. Nos experimentos do próximo capítulo, o valor 1000 foi utilizado para este parâmetro. Entre os 107.438 versos analisados no experimento principal, apenas 108 ou 0,1% deles atingiram este limite. Via de regra, portanto, a busca atinge seu objetivo de ser exaustiva. Deste modo, quando o algoritmo ordena os nós para escolher qual deve ser explorado a seguir, ele está se precavendo para o caso raro de o limite ser atingido: é melhor examinar as alternativas promissoras antes, pois não é impossível que a busca seja interrompida.

Quando o algoritmo tenta aplicar uma regra a uma elocução, ele passa à máquina de regras o parâmetro  $b_n$ , que especifica qual das ocorrências do fenômeno sendo tratado pela regra dentro da elocução deve ser considerada. Como o algoritmo não sabe quantas são ou mesmo se existem ocorrências a serem consideradas, ele começa com  $b_n = 1$  e progressivamente aumenta o parâmetro, até que a aplicação da regra falhe. O algoritmo adiciona um novo nó à árvore quando explora uma elocução existente, e por isso ele mantém registro de quais valores de  $b_n$  já foram utilizados para cada elocução e regra.

Mais do que explorar cada ocorrência individualmente, o algoritmo precisa gerar todas as combinações possíveis. Se uma regra pode ser aplicada em  $n$  posições diferentes de uma elocução, e contanto que a aplicação em uma ocorrência não elimine a possibilidade de aplicação em outra, existem  $2^n$  combinações a serem consideradas. Por esse motivo, a busca exaustiva atualmente empregada pelo gerador de elocuições seria um problema caso versos livres de 20, 30 ou mais sílabas fossem analisados: uma explosão combinatória estaria fadada a acontecer. Entretanto, em versos métricos, e portanto pequenos, o número de combinações fica baixo e o sistema não enfrenta maiores problemas.

As diferentes combinações de aplicação de uma regra a uma elocução são criadas naturalmente pelo algoritmo de busca. Mesmo que a partir de um nó os diferentes valores de  $b_n$  não gerem todas as combinações, as elocuições geradas a partir desse nó também são recursivamente exploradas pelo algoritmo, de modo que, se o limite máximo de elocuições não for atingido, todas as combinações serão eventualmente consideradas.

**As regras prosódicas.** À exceção das três regras que compõem as camadas 7 e 8, as demais regras são adicionadas ao sistema conforme a necessidade. Quanto mais regras o sistema possuir, menos regras precisam ser criadas quando um *corpus* novo é processado. Entretanto, a possibilidade de haver a necessidade de criar uma nova regra está sempre presente quando um *corpus* é processado pela primeira vez. O experimento principal do próximo capítulo, por exemplo, demandou 159 regras para processar 107.438 versos. Para averiguar o quão gerais são as regras criadas, esse mesmo experimento faz uma estimativa da quantidade de regras que precisam ser adicionados quando um novo *corpus* é processado.

A pontuação das regras foi determinada empiricamente com o auxílio de especialistas. Quanto mais gerais são as regras, mais difícil é atribuir-lhes uma pontuação que funcione bem em todos os contextos; a pontuação de regras com poucos casos de uso, por outro lado, pode ser estimada com mais facilidade. Um outro desafio para a pontuação reside na diversidade, tanto estilística quanto temporal, dos *corpora* do experimento: uma mesma regra pode ser vista como obrigatória por um poeta mas inteiramente opcional por outro. Em casos assim buscou-se, sempre que possível, um equilíbrio entre os extremos.

As regras, com todos seus parâmetros — inclusive a pontuação de penalidade —, constam do Apêndice B. Aqui serão apresentados exemplos representativos de regras para cada uma das camadas. Eles mostram a elocução antes e depois da aplicação da regra — mas o *depois* não representa necessariamente o estágio final; de fato, algumas regras, como a eclipse, existem justamente com o intuito de que outras regras venham na sequência.

**Camada 1.** Contém regras que tornam tônicas certas palavras átonas em fim de verso. Uma forma como *nas*, da junção da preposição *em* com o artigo *as*, é átona, mas pode ser feita tônica se o poeta o quiser, como neste exemplo:

ANJ.54.2.2 Do céu, em reflexos, nas

	1	2	3	4	5	6	7
<i>antes</i>	du	sew	ẽ	be	fle	jus	nas
<i>depois</i>	du	sew	ẽ	be	fle	jus	'nas

**Camada 2.** Contém regras que aplicam fenômenos que removem ou adicionam sons sem combiná-los, ou então alteram a posição do acento sem alterar a qualidade dos fonemas da elocução. Trata-se da camada com

maior variedade de regras. Os fenômenos incluídos nesta camada são:

- Aférese, a remoção de um fonema no início da palavra:

DUR.4.81.5 Acomete o esquadrão com fúria brava,

	1	2	3	4	5	6	7	8	9	10	11	12	13
antes	a	ko	me	ti	u	es	kwa	drāw	kō	fu	rja	bra	va
depois	a	ko	me	ti	us	kwa	drāw	kō	fu	rja	bra	va	

- Síncope, a remoção de um fonema no interior da palavra:

SIL:1.105.4.1 a se agrupar em cúmulos no horizonte,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
antes	a	si	a	gru	par	ē	ku	mu	lus	nu	o	ri	zō	ti
depois	a	si	a	gru	par	ē	ku	mlus	nu	o	ri	zō	ti	

- Apócope, a remoção de um fonema no fim da palavra:

MAT:38.21.1.46 deste mármore, deste roble,

	1	2	3	4	5	6	7	8	9
antes	des	ti	mar	mu	ri	des	ti	wo	bli
depois	des	ti	mar	mur	des	ti	wo	bli	

- Epêntese, a inserção de uma vogal de apoio no meio de um grupo consonantal:

SIL:1.59.3.1 Qual magnólia que à manhã descerra,

	1	2	3	4	5	6	7	8	9	10	11	12
antes	kwal	mag	no	lja	ki	a	ma	jā	de	se	wa	
depois	kwal	ma	gi	no	lja	ki	a	ma	jā	de	se	wa

- Paragoge, a inserção de uma vogal de apoio no fim de uma palavra:

PIN.27.15.3 Sob verde leão, sofre mesquinha.

	1	2	3	4	5	6	7	8	9	10	11
antes	sub	ver	di	le	āw	so	fri	mes	ki	jā	
depois	su	bi	ver	di	le	āw	so	fri	mes	ki	jā

- Sístole, o recuo do acento mais para o início da palavra:

SIL:1.57.4.5 nem harmonia mais rara

	1	2	3	4	5	6	7	8
antes	nē	ar	mo	ni	a	majs	wa	ra
depois	nē	ar	mo	ni	a	majs	wa	ra

- Diástole, o avanço do acento mais para o fim da palavra:

FIG:4.39.2.1 E teríamos gozado o que gozamos

	1	2	3	4	5	6	7	8	9	10	11	12	13
antes	i	te	ri	a	mus	go	za	du	u	ki	go	zā	mus
depois	i	te	ri	a	mus	go	za	du	u	ki	go	zā	mus

**Camada 3.** Contém regras que aumentam ou diminuem o tamanho de uma palavra, de dois tipos:

- Sinérese, a contração de vogais numa sílaba, convertendo pelo menos uma das vogais em semivogal:

MAT:25.9.1.10 para a gala coadjutores.

1 2 3 4 5 6 7 8 9 10  
*antes* 'pa:ra:a'ga:la: ko ad:zu'to:ris  
*depois* 'pa:ra:a'ga:la:kwad:zu'to:ris

- Diérese, a regra inversa, que desmancha um ditongo, convertendo uma semivogal em vogal:

CAM.3.96.7 Em constituições, leis e costumes,

1 2 3 4 5 6 7 8 9 10 11  
*antes* ẽ kōs | ti | tu | j | sōjs | | lejs | i | kos | tu | mis |  
*depois* ẽ kōs | ti | tu | i | sōjs | | lejs | i | kos | tu | mis |

**Camada 4.** Aqui estão regras que reduzem ou apagam vogais, para que nas camadas subsequentes elas sofram sinalefa:

FIG:4.20.4.3 Ou era o dos vermes que de mim zombavam!

1 2 3 4 5 6 7 8 9 10 11 12 13  
*antes* ow' ɛ ra u dus' ver mis ki di' mĩ zõ' ba vãw  
*depois* u ɛ ra u dus' ver mis ki di' mĩ zõ' ba vãw

**Camada 5.** Contém as regras de eclipse, que removem a trava nasal final de certas palavras, expondo a vogal final para que ela sofre uma subsequente sinalefa:

MAT:13.8.1.86 porque servem um falso antojo

1 2 3 4 5 6 7 8 9 10  
*antes* pur ki ser ṽẽ ã fal su ã to zu  
*depois* pur ki ser vi ã fal su ã to zu

**Camada 6.** As regras de elisão ficam nesta camada, removendo a vogal átona final /a/, possibilitando inclusive sinalefas complexas com as vogais seguintes:

COS.111.47.2 A quem, nada ignorando do que eu sinto,

1 2 3 4 5 6 7 8 9 10 11 12 13  
*antes* a | kē | na | da | ig | no | rã | du | du | kɪ | ew | sī | tu  
*depois* a | kē | na | dig | no | rã | du | du | kɪ | ew | sī | tu

**Camada 7.** Está aqui a regra especial que congela as sílabas internas, sem nenhum impacto visível.

**Camada 8.** Aqui ficam as duas regras fonológicas obrigatórias do português:

- A regra que torna um /s/ final sonoro quando a palavra que o segue começa com vogal ou consoante sonora:

PIN.55.26.2 Que vos prende e depois dá liberdade,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	kɪ	vʊs	prɛ	dɪ	ɪ	dɛ	pojs	da	li	ber	da	dɪ
<i>depois</i>	kɪ	vʊs	prɛ	dɪ	ɪ	dɛ	pojz	da	li	ber	da	dɪ

- A regra que reposiciona algumas consoantes finais quando a palavra seguinte começa com vogal:

COS.49.3.3 De um peito entre os heróis esclarecido;

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	dɪ	ũ	pej	tu	ẽ	trɪ	u	ze	rɔjz	es	kla	re	si	du
<i>depois</i>	dɪ	ũ	pej	tu	ẽ	trɪ	u	ze	rɔj	zes	kla	re	si	du

**Camada 9.** Contém as regras de crase, que fundem vogais parecidas numa só:

MAT.29.19.1.9 lhe rendi a alma mil vezes

	1	2	3	4	5	6	7	8	9
<i>antes</i>	kɪ	ʊẽ	di	a	al	ma	mil	ve	zɪs
<i>depois</i>	kɪ	ʊẽ	di	al	ma	mil	ve	zɪs	

**Camada 10.** As regras de sinalefa estão nesta camada, desde as mais simples às mais complexas:

MAG.16.6.3 E antes que a Humanidade entrever possa

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	jã	tɪs	kɪ	a	u	ma	ni	da	djẽ	tre	ver	pɔ	sa
<i>depois</i>	jã	tɪs	kjaw	ma	ni	da	djẽ	tre	ver	pɔ	sa		

## 5.8. Análise métrica

Esta etapa do sistema analisa o conjunto das elocuições encontradas num poema e propõe uma métrica para os seus versos. Mais especificamente, a análise métrica classifica o poema em uma das seguintes classes:

1. Classe I: o poema é isométrico. Todos os versos possuem exatamente a mesma quantidade de sílabas poéticas.
2. Classe D: o poema contém exatamente duas medidas. Este é o caso típico de poemas com versos quebrados, como os decassílabos e seus quebrados de seis sílabas, que são bastante comuns.
3. Classe E: o poema contém mais de dois metros, mas no máximo dois por estrofe. Portanto, cada estrofe é analisada separadamente e classificadas nas classes homônimas: I quando contiverem uma medida e D quando duas.
4. Classe P: o poema contém mais de duas medidas por estrofe, ou seja, é polimétrico.

Essas classes evidentemente não cobrem a totalidade dos poemas em português, mas abrangem boa parte deles. Casos não previstos são poemas com versos compostos, que são confundidos pelo sistema com poemas de classe I, poemas com vários metros distribuídos de forma simétrica em suas estrofes, que são confundidos com poemas de classe P, entre outros.

Além de categorizar o poema e estrofes em classes, o sistema atribui a cada verso um tamanho específico. Quando o poema é isométrico, a atribuição é trivial; por outro lado, no caso de poemas de classe D, P e das estrofes de classe D o sistema precisa decidir qual é o tamanho de cada verso individualmente.

O processo de classificação é sequencial e, uma vez que uma categoria aceitável tenha sido encontrada, ele é interrompido. Para decidir se o poema pertence à classe I, um teste de isometria é realizado. Se o teste falhar, um teste de dimetria é levado a cabo na tentativa de atribuir o poema à classe D. Se também esse falhar, o sistema percorre todas as estrofes e tenta classificá-las em I e D; se conseguir, o poema como um todo recebe o rótulo de E. Caso contrário, o poema é classificado como P sem nenhum teste.

**Teste para a classe I.** O teste de isometria pode ser realizado para um grupo de versos qualquer, de maneira que tanto os versos de um poema quanto os versos de uma estrofe podem ser submetidos ao teste. A ordem dos versos também não é importante. O teste é realizado da seguinte

forma: para cada tamanho métrico possível  $t$ , um índice de desvio  $d_t$  é calculado. Se este índice for menor que o limiar  $L_I$  para algum tamanho  $t$ , então o poema é classificado como isométrico com versos de tamanho  $t$ .

O desvio geral  $d_t$  é igual à média dos desvios de cada verso. Estes últimos são encontrados da seguinte maneira. É encontrada inicialmente a elocução com menor pontuação pts, ou seja, a elocução que o sistema julgou tem a maior naturalidade. Em seguida é buscada a elocução com menor pontuação de naturalidade entre aquelas que possuem  $t$  sílabas poéticas. A diferença absoluta de pontuação entre as duas é o desvio do verso; se as duas elocuições coincidirem, o desvio do verso é zero. Se não houver nenhuma elocução com  $t$  sílabas poéticas, o sistema considera que o desvio do verso é  $D$ , um parâmetro configurável do sistema. Com  $D = \infty$ , qualquer verso que não possuir elocuições com  $t$  sílabas poéticas inviabiliza este tamanho.

A faixa de tamanhos a serem testados é definida pelo tamanho mínimo e máximo encontrado entre todas as elocuições de todos os versos. Entre dois tamanhos  $t$  e  $q$  com  $d_t = d_q < L_I$ , é escolhido o menor entre  $t$  e  $q$ .

**Teste para a classe D.** Este teste é feito de forma similar ao de isometria. Para cada par  $t, q$  de tamanhos possíveis é calculado um índice de desvio  $d_{tq}$  — contanto que  $|t - q| \geq 2$ ; o sistema considera que um poema não pode conter duas medidas consecutivas, como 9 e 10 ou 10 e 11. Se houver pelo menos um par  $t, q$  com  $d_{tq} < L_D$ , em que  $L_D$  é um valor de limiar, então o poema é categorizado com a classe D e a cada verso é atribuído um dos tamanhos  $t$  ou  $q$ . Em caso de empate entre dois pares de tamanho cujos desvios estejam abaixo do limiar, o primeiro encontrado é utilizado.

O desvio  $d_{tq}$  depende novamente dos desvios de cada verso, que são calculados do mesmo modo utilizado pelo teste isométrico, à exceção de que para cada verso são calculados dois desvios, um para cada tamanho. O desvio final de um verso é o menor valor entre os dois, ou então  $D$ , caso o verso não possua nenhuma elocução com tamanho  $t$  ou  $q$ . O tamanho do qual se derivou o menor desvio é também atribuído como tamanho final do verso, caso o par  $t, q$  esteja abaixo do limiar  $L_D$ .

**Teste para a classe E.** Este teste utiliza os dois outros, para as classes I e D, a todas as estrofes do poema. O teste I é realizado primeiro e, se obtiver êxito, classifica os versos da estrofe correspondente; caso falhe, o teste D é similarmente empregado. Se ambos falharem, o teste para a classe E falha.

Se, por outro lado, todas estrofes puderem ser classificadas em I ou D, a média dos desvios de cada classificação é comparada com um limiar  $L_E$ ; se estiver abaixo deste limiar, o teste é exitoso.

**Os demais casos: a classe P.** Se todos os testes falharem, o poema é classificado como P, polimétrico. O sistema precisa, mesmo assim, estabelecer o tamanho provável dos versos do poema. Para fazê-lo o sistema computa, com base nas elocuições mais naturais de cada verso, a frequência dos diversos tamanhos silábicos, ou seja, encontra um *histograma silábico* para o conjunto de versos.

O histograma associa a cada tamanho silábico possível uma contagem, correspondente à quantidade de versos que naturalmente possuem aquele tamanho. Cada verso dispõe de uma unidade para a contagem do histograma; se houver apenas uma elocução com pontuação mínima para um dado verso, então o número de sílabas poéticas daquela elocução é incrementado no histograma. Se houver  $n > 1$  elocuições, cada uma soma  $1/n$  à posição correspondente do histograma. Por fim, o histograma é normalizado de acordo com a quantidade total de versos na qual ele é baseado, de maneira que a soma de todos os seus valores é 1 ou 100%.

O sistema ordena então os tamanhos silábicos possíveis de acordo com a sua contagem no histograma. Da maior para a menor frequência, o sistema coleta tamanhos silábicos até que um limite mínimo de  $P = 95\%$  seja atingido; a intenção é que, ao selecionar esses tamanhos silábicos, a ampla maioria dos versos pode ser naturalmente escandida, e os que não puderem sê-lo de forma plenamente natural provavelmente o poderão com um pouco menos de naturalidade. É possível, contudo, que algum verso de presença minoritário no histograma não possa ser escandido de nenhuma forma com os tamanhos selecionados; se existirem tais versos, o sistema adiciona à lista de tamanhos que está construindo a quantidade de sílabas poéticas da elocução mais natural de cada um desses versos. Com a lista de tamanhos permitidos finalizada, a cada verso é atribuído o tamanho com o qual sua elocução mais natural tem a menor distância.

**Versos fragmentários e com erros.** O algoritmo de análise métrica aqui descrito é sensível a erros textuais, especialmente àqueles presentes em poemas pequenos. Um soneto decassilábico, por exemplo, pode confundir o sistema caso um de seus versos não contenha elocuições com menos de 12 sílabas, já que existem apenas 14 versos em que a penalidade  $D$  atribuída ao verso pode se dissolver. É possível que o poema seja, por esse motivo,



categorizado como D. Pior do que isso, é possível que existam outros versos no poema que podem mais naturalmente serem pronunciados em 12 sílabas, apesar de poderem sê-lo em 10 também. A tais versos o teste para a classe D atribuirá o tamanho 12, e não 10.

Há duas razões para o gerador de elocuições não encontrar pelo menos uma variante com o tamanho correto. Primeiramente, ele pode não dispor das regras necessárias para que aquele número de sílabas seja atingido. Alternativamente, o verso pode conter erros ou estar fragmentário, o que acontece com certa frequência na prática, conforme expôs o Capítulo 4. Para o tratamento cuidadoso de um *corpus*, convém, portanto, que casos particulares de versos mal comportados sejam examinados: se faltam regras ao sistema, elas podem ser adicionadas; se o verso está incorreto, ele pode ser marcado como tal, o que o elimina da análise métrica e reduz a sua interferência negativa no processo.

## 5.9. Análise rítmica

Esta etapa visa a encontrar, entre as elocuições com tamanho silábico compatível com os resultados da análise métrica, aquela que é mais ritmicamente aceitável. Para o seguinte verso dodecassilábico de Augusto dos Anjos encontrado durante os experimentos do próximo capítulo, por exemplo, o sistema gerou uma série de elocuições possíveis de 12 sílabas, entre as quais:

ANJ.196.11.3 Compensaria toda a sua dor imensa

	1	2	3	4	5	6	7	8	9	10	11	12	13
4-6-8-10-12	kõ	pẽ	sa	'ri	a	'tõ	da	'su	a	'dõ	ri	mẽ	sa
4-6-10-12	kõ	pẽ	sa	'ri	a	'tõ	da	a	'swa	'dõ	ri	mẽ	sa
3-6-8-10-12	kõ	pẽ	'sa	ri	a	'tõ	da	'su	a	'dõ	ri	mẽ	sa
3-6-10-12	kõ	pẽ	'sa	ri	a	'tõ	da	a	'swa	'dõ	ri	mẽ	sa
3-5-8-10-12	kõ	pẽ	'sa	rja	'tõ	da	a	'su	a	'dõ	ri	mẽ	sa

Os números à esquerda de cada elocução correspondem às posições silábicas que o sistema considera acentuadas. O sistema precisa decidir qual delas é mais viável, levando em consideração sua naturalidade e seu ritmo.

**O catálogo de ritmos.** Um leitor humano com um pouco de experiência em poesia em língua portuguesa tem algumas expectativas ao se deparar com um poema isométrico decassílabo. Mais concretamente, ele espera encontrar certos padrões rítmicos com mais frequência do que outros. O decassílabo com a 6ª sílaba acentuada, chamado heroico, é particularmente

comum, e em menor medida é também bastante usual o sáfico, com a 4ª (e em geral também a 8ª) acentuadas.

O Aoidos, ao analisar um poema grande, como um dos cantos do *corpus* CAM, o menor dos quais contém 696 versos, pode encontrar sem auxílio externo tais padrões rítmicos. Porém, a análise de um soneto ou de uma quadra dificilmente pode revelar muito a respeito dos padrões comumente utilizados na poesia. Se o sistema escolher como proposta final de escansão a elocução mais natural que se encaixar na métrica, pode cometer um erro que um humano, com suas expectativas, não cometeria. No seguinte verso dodecassilábico, do *corpus* FIG, o sistema considera que a elocução mais natural deixa o acento na quinta sílaba:

FIG:5.86.3.2 Ao cruel desalento; e a tua alma chorar,

1 2 3 4 5 6 7 8 9 10 11 12 13  
2-5-8-9-12 aw|'krwew|de|za|'lẽ|tu|ja|'tu|'al|ma|so|'rar|

No entanto, em português um verso de 12 sílabas geralmente deve ter sua 6ª acentuada. Esta alternativa é encontrada pelo sistema, que não a considera a mais natural:

1 2 3 4 5 6 7 8 9 10 11 12 13  
3-6-8-9-12 aw|'kru|'ew|de|za|'lẽ|tw|ja|'tu|'al|ma|so|'rar|

Uma maneira de tentar amenizar o problema é fornecer ao sistema, junto com o poema a ser processado, um *catálogo de ritmos* contendo os padrões rítmicos que devem ser privilegiados.

Este catálogo associa a cada comprimento silábico uma lista de combinações de posições silábicas, e a cada combinação um fator que expressa o quão comum ela é. A soma dos fatores de um mesmo comprimento silábico não pode ultrapassar 1. Para os versos decassílabos, por exemplo, o catálogo poderia conter a informação de que há dois padrões privilegiados: um com a 6ª acentuada e outro com a 4ª. À primeira poderia ser atribuído o fator 0,8, e, à segunda, 0,2.

De posse do catálogo, o sistema percorre todas as elocuições com o tamanho correto de sílabas e lhes atribui uma pontuação rítmica. A soma da pontuação rítmica e da pontuação de naturalidade de cada elocução é utilizada como índice final para a seleção da proposta de escansão: a elocução com a menor soma é escolhida.

A pontuação rítmica de uma elocução é computada tentando-se encontrar uma entrada no catálogo de ritmos que seja compatível com a elocução. Uma elocução é *compatível* com o padrão definido no catálogo se todas as posições silábicas definidas nele são tônicas na elocução. A presença de sílabas tônicas na elocução além daquelas definidas no padrão não interferem na compatibilidade. O catálogo é percorrido de forma se-

quencial e a primeira compatibilidade encontrada é a utilizada, as demais sendo desprezadas; se nenhuma entrada for encontrada, à pontuação rítmica é atribuída o valor zero. Ademais, apenas entradas com fator acima de 0,001 são consideradas.

A pontuação rítmica  $p_c$  é diretamente calculada a partir do fator  $f$  do padrão compatível através da equação

$$p_c = -a \log_{10}(f) + b,$$

em que  $a$  e  $b$  são parâmetros cujos valores presentes são 3 e 10. O valor de  $p_c$  é negativo porque ele segue a pontuação de naturalidade, em que valores menores são melhores. A intenção dessa fórmula é permitir que mesmo fatores baixos possam ter alguma pontuação minimamente expressiva e evitar que fatores altos produzam pontuações rítmicas fortes demais; os valores de  $a$  e  $b$  foram escolhidos com essa intenção em mente. Um fator de 0,001, por exemplo, gera uma pontuação de -1,00, e um fator de 0,9 resulta em -9,86.

**Operação sem o catálogo.** O sistema pode operar sem o catálogo de ritmos. Neste caso, ele computa um histograma rítmico a partir da elocução mais natural de cada verso para em seguida calcular uma pontuação rítmica  $p_h$  que, somada à pontuação de naturalidade das elocuições, reordena as elocuições, potencialmente selecionando como proposta de escansão uma elocução menos natural porém mais aderente ao histograma. Em poemas com mais de um metro, ou seja, de classes D, E e P, o sistema agrupa todos os versos de um determinado comprimento e os analisa juntos.

O histograma rítmico de um conjunto de elocuições é computado a partir das sílabas acentuadas de cada uma. A cada posição silábica o histograma associa o percentual de elocuições em que a sílaba está acentuada. O poema em que se encontra aquele verso dodecassílabo, por exemplo, gerou o seguinte histograma rítmico:

$$\begin{array}{ccccccccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ \text{FIG:} & 5,86 & 0,29 & 0,43 & 0,64 & 0,21 & 0,14 & 0,93 & 0,07 & 0,57 & 0,43 & 0,14 & 0,07 & 1,00 & 0,00 \end{array}$$

A sexta sílaba é claramente privilegiada: 93% dos versos do soneto (13 dos 14) a tem acentuada em sua elocução mais natural; o único verso excepcional é justamente o de exemplo.

A pontuação rítmica  $p_h$  de uma elocução é calculada comparando-se o esquema rítmico de cada sílaba com a posição correspondente no his-

tograma, ou seja,

$$p_h = \sum_{i=1}^n |h_i - e_i|,$$

em que  $n$  é o número máximo de sílabas encontradas,  $h_i$  é o valor do histograma na posição  $i$  e  $e_i$  vale 1 se a elocução tem uma sílaba tônica na posição em  $i$  e 0 caso contrário. Vê-se que a pontuação  $p_h$  é sempre positiva, pois trata-se de um valor punitivo: desvios do histograma são punidos, não incentivados.

**Geração automática do catálogo.** O catálogo de ritmos pode ser gerado automaticamente pelo sistema, com a intenção de que um catálogo possa ser produzido com base num grande número de versos e então usado para processar poemas individuais.

O algoritmo de geração automática do catálogo utiliza a proposta final de escansão dos versos. Os versos são agrupados conforme seu número de sílabas poéticas, independentemente de poema ou poeta. Para cada grupo, o sistema encontra todos os esquemas rítmicos e as respectivas frequências relativas, e também calcula o histograma rítmico do grupo.

A lista de esquemas rítmicos de cada tamanho silábico é ordenada, dos mais aos menos frequentes, e cada esquema é avaliado pelo sistema. As posições que estão acentuadas no esquema rítmico são convertidas em átonas se o histograma rítmico do grupo naquela posição  $t$  estiver abaixo de um valor de limiar  $G_t$ . O valor do limiar varia conforme o tamanho silábico, sendo

$$G_t = a - bt,$$

em que  $a$  e  $b$  são constantes com valores atuais 0,7 e 0,03. O intuito é que o limiar  $G_t$  diminua conforme versos mais longos são considerados, partindo-se do entendimento de que os padrões rítmicos são mais relevantes para versos mais longos. Assim, em versos de quatro sílabas as posições serão consideradas tônicas apenas se  $h_4 > G_4 = 0,58$ ; mas os dodecassilábicos precisam apenas superar  $G_{12} = 0,35$ . A última sílaba tônica dos padrões rítmicos nunca é convertida em átona, já que por definição 100% delas são tônicas.

Com base neste esquema rítmico alterado, o catálogo em construção é consultado: não apenas o padrão rítmico pode já constar dele, mas o padrão candidato pode ser compatível com algum outro. Um padrão  $p$  é compatível com outro  $q$  se todas as sílabas tônicas especificadas por  $q$  são

tônicas também em *p*. Se o catálogo em construção contiver o padrão rítmico 6-10, então um novo padrão candidato 4-6-10 não é adicionado ao catálogo, por ser compatível com 6-10. O sistema também mantém no catálogo em construção a quantidade de versos em que cada padrão foi baseado, de forma que, quando um padrão candidato não é adicionado ao catálogo por já haver algum compatível, a contagem de versos do candidato é somada à do padrão existente.

Ao fim da lista de padrões de um grupo, o catálogo é reordenado segundo o número de versos associado a cada esquema. A consequência dessa ordenação é que alguns padrões podem ficar inalcançáveis a uma busca sequencial, como a realizada pela análise rítmico com catálogo: se o padrão 4-10 constar no catálogo antes de 4-8-10, então nenhuma elocução jamais será casada com este último. Por esse motivo, o catálogo é percorrido e os esquemas rítmicos compatíveis com algum outro com prioridade maior são eliminados e sua contagem de versos somada à do outro.

O experimento principal do próximo capítulo produziu um catálogo de ritmos, que pode ser visto na Tabela 5.4. Este catálogo não é sem críticas; os decassílabos, por exemplo, não possuem um esquema 4-10 para os sáficos, apenas 4-8-10, o que significa que, para o sistema, um padrão como 3-10, 4-10 ou 5-10 são igualmente aceitáveis.

## 5.10. Detalhes técnicos

O Aoidos está escrito inteiramente em Python 3. As dependências externas são a biblioteca `lxml`, utilizada para ler os arquivos XML, e `PyMySQL`, empregada opcionalmente para a comunicação com o banco de dados relacional MariaDB. O programa opera em modo de lote: recebe um arquivo XML e o resultado é escrito na tela ou guardado em banco de dados. É possível também especificar qual poema, estrofe ou verso o sistema deve processar, utilizando uma notação específica de endereço. Cada poema é processado em paralelo pelo Aoidos, o que acelera bastante o processamento de grandes *corpora* quando há processadores disponíveis.

Existem vários níveis de verbosidade, mas, na configuração padrão, as informações geradas pelo sistema ao longo de sua execução incluem:

- Mapa de tamanhos das estrofes do poema, que pode ser usado para detectar erros no material fonte.
- Detalhes sobre a análise métrica, com tabelas de desvios para os testes das classes I, D e E e o histograma da classe P.

Tabela 5.4.: Catálogo de ritmos gerados pelo experimento principal do próximo capítulo.

Tamanho	Esquema	Fator
5	2, 5	0,8072
6	2, 6	0,5888
9	4, 9	0,6655
	6, 9	0,2513
10	6, 10	0,7995
	4, 8, 10	0,1826
11	2, 5, 8, 11	0,4410
	5, 11	0,3938
12	2, 6, 8, 12	0,2847
	3, 6, 8, 12	0,1849
	3, 6, 12	0,1829
	2, 6, 12	0,1410
	4, 6, 8, 12	0,0859
	4, 6, 12	0,0619

- Resultado da análise métrica, incluindo a classificação e o tamanho atribuído a cada verso do poema.
- Pontuação de naturalidade das elocuções após a análise métrica, bem como as pontuações adicionais da análise rítmica,  $p_c$  e  $p_h$ .
- Um mapa do esquema rítmico de cada verso do poema, o que possibilita a identificação visual de alguns padrões.
- A transcrição fonética dos versos, a divisão do texto original em suas sílabas e o esquema rítmico de cada um deles.
- A pontuação de naturalidade final de cada verso.
- As regras aplicadas a cada verso para a geração da escansão final — mas não todas as regras utilizadas durante a busca pelo gerador de elocuções.
- O número de elocuções encontradas pelo gerador de elocuções para cada verso.

### 5.11. Considerações metodológicas

O Aoidos foi desenvolvido com o propósito inteiramente prático de escandir versos. Para alcançar esse propósito, as etapas do sistema foram sendo gradualmente estabelecidas e refinadas. Algumas características do sistema estiveram bem definidas desde o início, enquanto outras se modificaram em face de problemas encontrados. Desde o princípio, por exemplo, tomou-se a decisão de se manipular fonemas e não grafemas; por outro lado, a interação entre as análises métricas e rítmicas tardou a tomar a forma atual.

O sistema foi pensado de forma a não necessitar de um conjunto de treinamento. Essa decisão foi tomada para evitar os custos de estabelecer um tal conjunto. Seria incorreto, contudo, afirmar que o estado atual do sistema não dependeu dos poemas utilizados ao longo do desenvolvimento das diversas etapas. O sistema não foi modelado e programado no vácuo; vários poemas ajudaram a estabelecer a estrutura e o comportamento de sistema. Alguns aspectos do sistema estão mais presos do que outros aos poemas usados quando de seu desenvolvimento. A etapa de localização do acento tônico de vocábulos, por exemplo, é completamente independente dos poemas; as regras prosódicas utilizadas pelo gerador de elocuições, porém, são bastante dependentes dos poemas em que foram baseadas, de forma que seria pouco surpreendente se fosse preciso criar novas regras prosódicas para tratar um *corpus* novo.

As etapas do sistema que aspiram a não sofrer modificações conforme novos *corpora* sejam tratados são as iniciais: desde o pré-processamento até a transcrição fonética. A etapa que mais precisa ser adaptada quando novos poemas são considerados é sem dúvida a geração de elocuições, em que residem as regras prosódicas; por esse motivo, o experimento principal de validação, descrito no próximo capítulo, busca estimar quantas regras novas são necessárias quando um novo *corpus* é tratado. As duas etapas restantes, de análise métrica e rítmica, podem sofrer adaptações no futuro, mas nesse caso provavelmente se trataria mais de enfrentar limitações — como é o caso dos versos compostos, que não são tratados pelo sistema atualmente — e menos de atacar particularidades do novo *corpus*.

Os trabalhos relacionados discutidos no Capítulo 3 forneceram inspiração para a modelagem do Aoidos, mas não o influenciaram diretamente. Características comuns entre os sistemas explicam-se pela natureza comum do problema tratado: se vários sistemas usam transcrição fonética,

a razão é o fato de que a poesia métrica depende da oralidade para existir. Além daqueles trabalhos relacionados, convém ainda mencionar outros que, embora não tratem da escansão automática de versos, possuem alguma relação com as etapas iniciais do sistema.

Os sistemas de síntese de fala tomam um texto escrito e o convertem em sons que imitam a voz humana. Tais sistemas tipicamente convertem o texto da representação ortográfica para a fonética como uma das etapas para a síntese da voz. Essa conversão frequentemente implica também a determinação da vogal tônica e a divisão silábica dos vocábulos. Ademais, é desejável que os sistemas falem de forma natural; para tanto, eles incorporam regras de prosódia para que a concatenação das palavras seja feita de modo harmônico. Existe, portanto, um terreno comum percorrido pelos sistemas de síntese de fala e de escansão automática. Apesar de o Aoidos ter sido desenvolvido de forma completamente independente de outros sistemas, cabe aqui destacar, sem pretensões de completude, alguns trabalhos anteriores de síntese de fala.

Oliveira, Viana e Trancoso (140) apresentaram o DIXI, um sistema de síntese de voz para o português europeu. O sistema utiliza 18 regras para localizar o acento das palavras e 200 para fazer a transcrição fonética; exceções a ambas classes de regras podem ser incluídas num dicionário. Entre as regras de transcrição fonética, encontram-se regras de sândi, que harmonizam a junção das palavras. Amaral, Carvalho, Caseiro, Trancoso e Oliveira (38) utilizaram mais tarde o DIXI como base para um sistema cujo objetivo é auxiliar na transcrição fonética de *corpora* falados. Para fazê-lo, o sistema parte da transcrição ortográfica do *corpus* e a convertem para uma representação fonética. Com a intenção de produzir uma transcrição fonética que tenha mais chances de corresponder à pronúncia do falante, o sistema aplica regras de sândi que vão além daquelas do DIXI original: não apenas as regras deste novo sistema levam em consideração fenômenos não tratados pelo DIXI (como a síncope em palavras como *para*), mas também as regras são aplicadas de forma a produzir várias alternativas possíveis, de forma semelhante ao que o Aoidos faz. O sistema DIXI também foi utilizado pelos sistemas de escansão LuCas e SAEP, descritos no Capítulo 3.

Para o português europeu, há ainda outros sistemas que encontram a vogal tônica, fazem a separação silábica e realizam a transcrição fonética. Teixeira, Silva Freitas, Gouveia, Olaszy e Németh (168) adaptaram o MULTIVOX, um sistema de síntese de voz previamente existente, para o português europeu, definindo, para tanto, regras de localização do acento



e de conversão de grafemas em fonemas. Mais tarde, Gouveia, Teixeira e Silva Freitas (101) estabeleceram ainda algoritmos de divisão silábica para o MULTIVOX. Barros e Weiss (49) propuseram um sistema de síntese de voz que adota um tratamento probabilístico para as etapas de transcrição fonética, localização do acento e separação silábica. Braga e Resende Jr (66) descreveram módulos baseados em regra para as três etapas, avaliando o desempenho de cada módulo de forma independente.

Egashira (88) criou um sistema de síntese de voz para o português brasileiro. Na etapa de transcrição fonética, o sistema utiliza um conjunto de regras que mapeiam grafemas a uma representação fonética. Entre as dificuldades dessa etapa, o autor destaca dois pontos: determinar a natureza aberta ou fechada dos grafemas *e* e *o* em posição tônica e descobrir o valor fonético do grafema *x*. O sistema em seguida encontra a vogal tônica da palavra utilizando 6 regras. Com base na transcrição fonética e na posição do acento, o sistema faz, finalmente, a divisão silábica da palavra.

Há ainda outros trabalhos realizados para o português brasileiro. Gomes (100) descreveu um sistema de síntese de voz que efetua a divisão silábica, determina a sílaba tônica e faz uma transcrição fonética dos vocábulos. Couto, Neto, Tadaiesky, Klautau e Maia (78) utilizaram o *framework* MARY para construir um sistema de síntese de voz; as etapas de transcrição fonética, divisão silábica e localização do acento são baseadas em regras, mas outras etapas do sistema utilizam modelos ocultos de Markov. Mais recentemente, Neto, Rocha e Sousa (138) descreveram e avaliaram separadamente a etapa de divisão silábica do sistema. Latsch (121) descreveu um sistema de síntese de voz em que a transcrição fonética é realizada com uma detecção preliminar do acento; a divisão silábica é então feita inteiramente com base nos fonemas; por fim, a localização definitiva do acento é feita.



## 6. EXPERIMENTOS

1	2	3	4	5	6	7	8	9	10	11
Me-	lbo-	r é ex	pri-	men-	tá-	lo	que	jul-	gá-	lo,
Mas	jul-	gue-o	quem	não	po-	de ex-	pri-	men-	tá-	lo.

Camões

Este capítulo descreve três experimentos de validação e dois casos de uso. Um experimento de validação é aplicado a uma etapa específica do sistema, a localização da vogal tônica, e os outros dois são voltados ao resultado final, as propostas de escansão. Dentre esses dois últimos experimentos, o primeiro e principal avalia a acurácia do sistema utilizando uma pequena proporção de um grande número de versos e o segundo o faz com uma grande proporção de um número reduzido de versos. Os dois casos de uso buscaram reproduzir dados estatísticos encontrados num livro (74) e num artigo (73).

A etapa de localização da vogal tônica é avaliada separadamente pois, além de ser de importância fundamental para a escansão, pode vir a ser usada em outras aplicações. A divisão silábica e a transcrição fonética são também importantes para a escansão e certamente úteis em outros domínios, mas não são tão críticas quanto a determinação da vogal tônica: uma divisão silábica que hesite entre *ca-i-rão* e *cai-rão* pode ser refeita pelo sistema através de sinéreses e diéreses; a natureza aberta ou fechada do grafema *e* em *sede* não afeta a escansão; mas o sistema não pode errar no posicionamento da vogal tônica em *avaro* ou *caiais*, sob pena de irremediavelmente comprometer a escansão, já que os metaplasmos de sístole e diástole não são comuns o suficiente para corrigir de forma efetiva possíveis erros do sistema na localização da vogal tônica. A avaliação das demais etapas, finalmente, se confundem com a avaliação do resultado final em si, de modo que não são avaliadas separadamente.

### 6.1. Vogal tônica

O desempenho do Aoidos com relação à detecção da vogal tônica (ou da atonicidade) das palavras foi avaliado através da verificação manual dos

*corpora* HOU10 e LIT30. No caso das palavras com hífen, o resultado produzido pelo sistema foi considerado correto somente quando todos os componentes da palavra tiveram sua vogal tônica corretamente identificada. Os erros correspondem a duas situações distintas. De um lado, o sistema desiste de encontrar a vogal tônica quando suas expectativas com relação às regras ortográficas da língua portuguesa não são atendidas, como é o caso de *darwinista* e *ginseng*, ambas do HOU10: a primeira palavra contém o grafema *w*, que não é usado em palavras nativas, enquanto a segunda apresenta o grafema *g* em posição final, o que não acontece com as palavras na nossa língua. A segunda situação que gera um erro nas análises desta seção é a identificação incorreta da posição da vogal tônica, como em *ununtrium* e *Be*, também do HOU10: o sistema acredita que a primeira palavra é oxitona (*ununtrium*) e que a segunda é uma palavra átona, quando a primeira, uma palavra não aportuguesada, deveria ser pronunciada *ununtrium*, e a segunda é um símbolo que deveria ter sido expandido por alguma etapa anterior do sistema, ora inexistente, o que resultaria na palavra *berllio*, cuja vogal tônica seria facilmente identificada pelo sistema.

O sumário da análise do *corpus* HOU10 consta da Tabela 6.1. A taxa de erros de 0,69%, apesar de baixa, merece alguns comentários. Dos 69 erros, 37 foram símbolos, abreviações ou siglas (agrupados na tabela sob a categoria Símbolos) que não estavam escritos por extenso, tornando nulos os esforços do sistema. Exemplos são *UV-A* (a radiação ultravioleta), *csch<sup>-1</sup>* (arco cossecante hiperbólica), *Pb* (o elemento chumbo) e *d* (a quarta letra do alfabeto). Os itens desta categoria só poderiam ser corretamente processados se estivessem escritos por extenso. A segunda maior fonte de erros foram as palavras técnicas, que em geral contém letras estranhas àquelas comumente utilizadas em português, como *wattado* (“electricamente ativo”) ou apresentam letras comuns em posições inesperadas, como *farad* (“unidade de capacitância”). A terceira fonte de erro são sufixos que o Houaiss inclui entre os verbetes, como *-ível* (presente em palavras como *imóvel*). A quarta fonte de erro são palavras com hífen que contém um prefixo que não pode ser considerado de forma independente, como *inter-relacionamento*. O prefixo, *inter-*, tem uma acentuação bastante independente, mas a ortografia não reflete essa independência, de forma que o sistema o pronuncia *inter*. Para que o Aoidos corretamente pronunciasse o prefixo, ele deveria estar escrito *ínter*, o que faz sentido se considerarmos que essa regra geral funciona para outras 860 palavras com hífen do *corpus* HOU10. A categoria de gentílicos contém as palavras *piquet-carneirense* (“relativo a Piquet Carneiro - CE”), cujo primeiro elemento tem uma ter-

Tabela 6.1.: Desempenho do Aoidos na detecção das vogais tônicas no corpus HOU10.

Acertos	9.931	99,31%
Erros	69	0,69%
Símbolos	37	0,37%
Técnicas	24	0,24%
Sufixos	5	0,05%
Prefixadas	3	0,03%
Gentílicos	2	0,02%
Onomatopeias	2	0,02%
Estrangeirismos	1	0,01%
Total	10.000	100,00%

minação anormal, e *seychellense* (“relativo à República de Seychelles”), que contém a letra *y*. A categoria de onomatopeias contém as palavras *vapt-vupt* e *vupt*, com terminações também anômalas. Por fim, a única palavra na última categoria foi *ginseng*, terminada por *g*.

Entre os erros apontados pela Tabela 6.1, os mais graves são aqueles das palavras prefixadas e das onomatopeias, pois são palavras nativas e corretamente grafadas no *corpus*. Os demais itens poderiam ser tratados por alguma outra parte do sistema ou então poderiam ter uma expensão ou aportuguesamento indicado no arquivo TEI.

Uma avaliação similar foi feita com todas as palavras do *corpus* LIT30, cujo resultado está sumarizado na Tabela 6.2. A taxa de acertos de 99,27% é bastante encorajadora, mas é preciso examinar os erros para conhecer as limitações do sistema.

Uma das maiores fontes de erro da localização da sílaba tônica em textos não tratados são fragmentos que não estão escritos por extenso. Um total de 53 numerais foram encontrados no texto; exemplos são 1859, *iv*, 6.<sup>a</sup> e 39º. Apenas uma sigla estava presente no *corpus*: trata-se do *CNPq*, que aparece nas informações sobre as edições digitais. Há ainda vários pronomes de tratamento, como *Dr.*, *Sr.*, *Ex.<sup>a</sup>* e até mesmo *Mme*.

Outra categoria importante de erros são as palavras encontradas em meio a texto em língua estrangeira, seja porque o autor está citando uma frase em outra língua, seja porque o autor compôs uma parte de sua obra em idioma diferente do português. O sistema não falhou em encontrar

uma pronúncia para todas as palavras estrangeiras; várias delas inclusive poderiam ser confundidas com palavras em português como *alma* (*alma* em italiano) e *for* (*por* em inglês). Mesmo no caso em que o sistema identificou corretamente a localização do acento na palavra estrangeira, o resultado foi considerado errado. Por mais que tenha acerto no acento de *questo* (*este* em italiano) e *juger* (*julgar* em francês), o sistema acertou por mera coincidência.

Entre os erros cometidos em palavras escritas por extenso num texto em português é possível identificar várias categorias. Existem as palavras que não foram inteiramente aportuguesadas: há aí galicismos, como *toilette* (hoje *toaleta*), latinismos, como *fac* de *fac totum* e até mesmo italianismos, como *Signora*. Os nomes próprios escritos com grafias estrangeiras representam um caso mais complexo, pois não são sempre passíveis de aportuguesamento. Exemplos são *Goethe*, *Leconte* e *Cavendish*. Novamente, tais palavras foram consideradas incorretas independentemente do resultado produzido pelo sistema.

Em um total de 18 casos, o sistema acertou na localização do acento considerando-se o que estava escrito, porém, em termos práticos, estava errado. Nesta categoria encontram-se palavras que estavam escritas incorretamente no texto, o que impediu que o sistema detectasse a real posição do acento. Por exemplo, duas vezes ocorre a palavra *porem* (que, conforme foi averiguado no texto, *não* é uma flexão do verbo *pôr*), que o sistema acentuou *porem*, o que não corresponde com a intenção do texto original, que era dizer *porém*. Também foram considerados erros quatro casos em que o programa acertou o acento apesar de a palavra não ter sido escrita corretamente: são as formas *Impróvida* (que deveria começar com *i*), *PEN-DENCIA* (que deveria ter um *I* no lugar do *l*), *esperá* (que ficou separada por espaço do complemento *la*) e *querelo* (em vez de *querê-lo*). Na análise tais casos foram considerados incorretos porque, como anteriormente, o sistema não sabe o que faz.

Por fim, o sistema não conseguiu determinar a posição do acento para mais três palavras: uma com ortografia que não foi atualizada, *Mayas* (em vez de *maias*), *d11-Rei* (em vez de *d'El-Rei*) e *mac-lama*, cuja interpretação não pôde ser estabelecida pelo autor da presente.

A taxa de acerto obtida pelo Aoidos nesta etapa é comparável àquela relatada na literatura. Braga e Resende Jr (66), por exemplo, propuseram um sistema capaz de localizar a vogal tônica corretamente em 99,54% dos casos. Os autores, contudo, excluem “siglas, acrônimos, numerais, etc.”; se aplicarmos os mesmos critérios aos números das Tabelas 6.1 e 6.2, as taxas

Tabela 6.2.: Erros e acertos feitos pelo sistema no *corpus* LIT30.

Acertos		29.780	99.27%
Erros		220	0.73%
Abreviações		32	0.11%
Números		53	0.18%
Texto estrangeiro			
	Italiano	31	0.10%
	Francês	31	0.10%
	Espanhol	6	0.02%
	Inglês	6	0.02%
	Latim	7	0.02%
	Alemão	1	0.00%
Texto português			
	Grafia estrangeira	8	0.03%
	Nome próprio estrangeiro	20	0.07%
	Erro <i>porem</i>	18	0.06%
	Erro <i>querelo</i>	4	0.01%
	Ortografia incorreta	1	0.00%
	Falha do sistema	1	0.00%
	Outros	1	0.00%
Total		30.000	100.00%

de acerto do Aoidos sobem para 99,73% para o *corpus* HOU10 e 99,55% para o LIT30.

## 6.2. Experimento principal de validação

Este experimento consistiu na aplicação do Aoidos a pouco mais de 100 mil versos e na consequente validação dos resultados produzidos. Os critérios de seleção, revisão e marcação dos 10 *corpora* utilizados no experimento encontram-se pormenorizados na no Capítulo 4.

### 6.2.1. Execução

O Aoidos foi executado numa instância do serviço Amazon Elastic Compute Cloud, de tipo *c4.2xlarge*, que dispõe de 8 núcleos de processamento e 16 GB de memória RAM. O Aoidos processa um arquivo TEI por vez e é capaz de analisar cada poema dentro do arquivo em paralelo, de forma que os núcleos de processamento são utilizados efetivamente pelo sistema.

Tabela 6.3.: Tempo de execução do sistema ao analisar os *corpora*.

<i>Corpus</i>	Versos	Tempo	Versos/s
CAM	8.816	51	172,9
MAT	33.840	127	<b>266,5</b>
DUR	6.672	42	158,9
COS	7.682	35	219,5
GON	4.184	20	209,2
MAG	5.637	32	176,2
PIN	14.233	81	175,7
ANJ	6.594	67	98,4
FIG	8.233	78	105,6
SIL	11.547	49	235,7
	107.438	582	184,6

A quantidade de memória mostrou-se bem além das necessidades do sistema.

Para a realização do experimento, o Aoidos foi executado duas vezes: uma para a geração do catálogo de ritmos e outra já de posse dele. A execução sem o catálogo levou 9 minutos e 59 segundos; com ele, 9 minutos e 42 segundos. Este último tempo de execução encontra-se detalhado na Tabela 6.3. Há duas razões principais para a taxa de versos por segundo variar conforme o *corpus*. Primeiramente, versos maiores demandam mais tempo. Eles contêm mais grafemas, mais sílabas e mais encontros vocálicos, o que faz com que um número maior de elocuições seja considerado para escandi-los. Segundamente, existem características de cada poeta e obra que fazem com que haja mais ou menos maneiras de se ler um dado verso. Por exemplo, os grandes poemas decassilábicos, CAM, DUR, GON e PIN, apesar de terem o mesmo número de sílabas, veem sua taxa de versos por segundo variar de 158,9 a 209,2 versos por segundo.

Foram escolhidos, de forma aleatória e indistinta entre todos os *corpora*, 1080 versos, ou seja, 1% do total, para serem examinados manualmente; todos eles constam do Apêndice C. Os versos foram escandidos pelo Aoidos no contexto do poema em que se encontram, com a utilização do catálogo de ritmos. Os versos foram considerados no seu contexto, para que o ritmo e sobretudo a métrica fosse compatível com os demais versos do poema, o que é de importância fundamental no caso dos poemas polimétricos. Os versos foram avaliados com base na transcrição fonética final gerada pelo sistema, na divisão do texto original em suas sí-



labas e no texto original. Assim, as três seguintes representações estavam disponíveis durante o exame:

ANJ.12.75.1 As pálpebras inchadas na vigília,

1	2	3	4	5	6	7	8	9	10	11
As	pál	pe	bra	s in	cha	das	na	vi	gí	lia,
as	'pal	pi	bra	zĩ	'ʃa	daz	na	vi	'zi	lja

Dois avaliadores trabalharam em cooperação na avaliação manual: o autor da presente e um especialista. As escansões produzidas pelo sistema foram classificadas em corretas ou incorretas. A pergunta utilizada para decidir se a escansão estava correta ou não foi a seguinte: *Existe uma escansão melhor que a produzida pelo sistema?* Se existe, então o sistema errou. Em alguns versos, o que é melhor ou não é bastante subjetivo; nestes casos a questão foi decidida de acordo com a experiência do especialista.

Versos parcial ou inteiramente escritos em língua estrangeira sem nenhuma marcação feita durante a revisão pontual foram considerados corretos se o sistema produziu uma escansão correta, incluindo a divisão silábica e a localização dos acentos. O verso seguinte, por exemplo, embora escrito em espanhol, foi bem escandido pelo sistema e portanto considerado correto:

MAT:6.15.7.1 A los Moros por dinero

1	2	3	4	5	6	7	8
A	los	Mo	ros	por	di	ne	ro
a	luz	mo	rus	pur	di	ne	ru

Por outro lado, o verso seguinte foi considerado incorreto, já que o sistema não soube nem identificar a sílaba tônica de *mihi*, nem dividi-la corretamente em sílabas:

MAT:6.3.1.76 de um parce mihi cantado.

1	2	3	4	5	6	7	8
de	um	par	ce	mi	h	can	ta
						do	
djũ	'par	sĩ	mi	'i	kā	'ta	du

A qualidade da transcrição fonética em si não foi levada em consideração. Em outras palavras, os versos não foram considerados incorretos se o programa pronunciou os versos como um sotaque estrangeiro. Por outro lado, não foram ignorados aspectos que dependem direta ou indiretamente da transcrição fonética, como a correta divisão silábica e a colocação das sílabas certas nos lugares certos.

O ritmo do verso foi avaliado durante o experimento no que tange à distribuição das sílabas. Não foi a intenção deste experimento avaliar se o sistema soube encontrar uma acentuação para o verso como um todo.

Considere-se o seguinte verso:

COS.1.1.2 Tomo entre vós, ó montes, o instrumento;

	1	2	3	4	5	6	7	8	9	10	11
	To-	mo	en-	tre	vó-	s,	ó	mon-	te-	s,	o ins-
											tru-
											men-
											to;
1-2-4-5-6-10	'to	'mwẽ	tri	'vɔ	'zɔ	'mõ	ti	zwĩs	tru	'mẽ	tu

As palavras *vós*, *ó* e *montes* são todas tônicas, mas não possuem todas a mesma importância nesse verso; a preposição *entre* tem um acento na primeira sílaba, mas ele é muito fraco para ter qualquer importância nesse verso. Em particular, o esquema rítmico do verso certamente não seria descrito por um especialista como 1-2-4-5-6-10; com a palavra *ó* anulando-se acentualmente entre duas tônicas e especialmente diante da 6ª, e com *entre* tornando-se plenamente átona, o esquema rítmico seria um muito mais plausível 1-4-6-10. Entretanto, por ter distribuído as sílabas corretas nas posições corretas, o verso seria considerado correto.

Seria mais natural e prudente que os 1080 versos fossem escandidos manualmente e que a resposta do sistema fosse cotejada à escansão manual. Fazê-lo, entretanto, acarretaria uma série de complicações. Em primeiro lugar, o custo de dividir as manualmente as sílabas de 1080 versos e codificá-las num banco de dados é muito maior do que examinar visualmente as mesmas divisões. Em segundo lugar, um tal banco de dados precisaria conter não apenas uma, mais várias alternativas de escansões aceitáveis, o que multiplicaria o trabalho de escansão manual. Em terceiro lugar, não está claro o quão fácil seria enumerar as escansões aceitáveis; é possível que alternativas aceitáveis de escansão escapem à atenção do especialista. Em quarto lugar, finalmente, para averiguar se o sistema produziu escansões corretas através dos metaplasmos corretos, o banco de dados precisaria codificar também os metaplasmos que afetaram cada escansão manual, o que aumentaria ainda mais o custo de construção do banco de dados. Por esses motivos, a avaliação foi feita verificando-se manualmente as escansões propostas pelo sistema.

### 6.2.2. Resultados

A Tabela 6.4 sumariza os resultados da validação manual. A taxa de acerto do sistema variou entre 97,5 e 100,0%, conforme o *corpus*. O percentual de versos escolhidos não foi constante entre os *corpora*, variando entre 0,9 e 1,1%, porque os versos foram selecionados aleatoriamente entre todos os *corpora*.

A taxa de acerto global do Aoidos foi de 99,0%, bastante acima dos 82,2% obtidos pelo SAEP, a proposta mais concreta de escansão automática

Tabela 6.4.: Resultados da verificação manual.

Corpus	Selecionados		Acertos		Erros	
	Total	%	Total	%	Total	%
CAM	87	1,0	87	<b>100,0</b>		
MAT	358	1,1	355	99,2	3	0,8
DUR	63	0,9	62	98,4	1	1,6
COS	81	1,1	79	97,5	2	<b>2,5</b>
GON	38	0,9	38	<b>100,0</b>		
MAG	62	1,1	61	98,4	1	1,6
PIN	141	1,0	139	98,6	2	1,4
ANJ	65	1,0	64	98,5	1	1,5
FIG	78	0,9	77	98,7	1	1,3
SIL	107	0,9	107	<b>100,0</b>		
	1.080	1,0	1.069	99,0	11	1,0

para o português antes do Aoidos. Esta diferença de 16,8 pontos percentuais é uma diferença *mínima*, já que a avaliação do SAEP levou em conta apenas o número de sílabas poéticas encontradas para os versos, sem considerar a escansão. É importante observar, contudo, que o SAEP não tem como objetivo primário a escansão de versos; de fato, a própria palavra *escansão* não aparece nenhuma vez na dissertação que descreve o sistema (124). A comparação das taxas de acerto, portanto, precisa ser feita tendo-se em mente a natureza diferente dos sistemas.

**Versos corretos com peculiaridades.** Entre os 1.069 versos corretamente escandidos pelo sistema destacaram-se 6 que apresentam certas particularidades. (1) Um deles, mencionado anteriormente, é um verso em espanhol, mas que, devido às similaridades fonéticas e ortográficas entre aquele idioma e o português, o sistema mesmo assim conseguiu escandir:

MAT:6.15.7.1 A los Moros por dinero

1 2 3 4 5 6 7 8  
A | los | Mo- | ros | por | di- | ne- | ro  
a | luz | 'mɔ | rus | pur | di | 'nɛ | ru

(2) Um verso havia sido marcado inescandível durante a revisão pontual, por ter 11 sílabas num poema decassilábico:

MAT:15.12.3.2 Que esse lhe dá de render o pão da me-,

1 2 3 4 5 6 7 8 9 10 11  
Que e- | sse | lhe | dá | de | ren- | de- | r o | pão | da | mê-  
'kʲɛ | sɪ | ʎɪ | 'da | dɪ | bẽ | 'dɛ | ru | 'pãw | da | 'me

O verso está escrito incorretamente na edição de referência deste *corpus*. Conforme informa Chociay (74), o texto correto é o seguinte, perfeitamente escandível em 10 sílabas:

\*Que esse lhe há de render o pão da me-,

Este verso foi, portanto, classificado como corretamente escandido.

Há o caso de 2 versos que também estavam textualmente incorretos no *corpora*, embora a revisão pontual não os houvesse detectado porque uma escansão foi encontrada pelo sistema dentro do número de sílabas esperado. (3) O primeiro é um verso de Camões:

CAM.5.62.2 Posto que todos Etiopes eram,

1	2	3	4	5	6	7	8	9	10	11	
Pos-	to	que	to-	do-	s	E-	tí-	o-	pe-	s e-	ram,
'pos	tu	ki	'tɔ	du	ze	'ti	u	pi	'ze	rãw	

Este verso tem um questionável esquema rítmico 1-4-7-10; de fato, a edição de referência esclarece que a palavra *Etiopes* deve ser grafada sem o diacrítico, *Etiopes*, o que resulta no esquema 1-4-8-10, um sáfico. (4) No outro verso, o problema na edição digital é a *falta* de um diacrítico:

COS.146.2.6 Bem que nesta copia o pinto,

1	2	3	4	5	6	7	8	
Bem	que	nes-	ta	co-	pia	o	pin-	to,
'bē	ki	'nes	ta	ko	'piw	'pī	tu	

Na edição de referência, a palavra *copia* está grafada *cópia*.

Os dois versos restantes pertencem a um mesmo poema, cujas estrofes contêm quatro versos: três heptassílabos e um que ora é tetrassílabo, ora trissílabo. Segundo Chociay (74), aparecem em poemas assim fenômenos de ligação entre versos, que regularizariam a maioria dos últimos versos de cada estrofe, deixando-os com três sílabas. (5) O primeiro verso do poema em questão é o seguinte:

MAT:36.2.9.4 sendo estrangeiros!

1	2	3	4	5	
sen-	do	es-	tran-	gei-	ros!
'sē	dwes	trã	'zej	rus	

Neste caso, o fenômeno de regularização aplicável seria a *compensação*: o verso anterior é agudo e portanto acomodaria uma sílaba a mais, ou seja, a primeira de *sendo*, ficando o verso à mão com apenas três sílabas. (6) O segundo verso é este:

MAT:36.2.42.4 a sua prata.

1	2	3	4	5
a	su-	a	pra-	ta.
a	'su	a	'pra	ta

Neste segundo caso, o verso se ligaria ao precedente por meio da *sinafia*: uma sinalefa entre a vogal final de um e a inicial de outro, deixando também esse com três sílabas.

**Versos incorretos.** Os 11 versos que o sistema escandiu incorretamente são aqui discutidos. (1) O primeiro já foi apontado anteriormente, trata-se de um verso de Gregório de Matos que possui o trecho em latim *parce mihi*:

MAT:6.3.1.76 de um parce mihi cantado.

1	2	3	4	5	6	7	8
de	um	par	ce	mih	i	can	ta
						do	
djũ	par	sĩ	mi	i	kã	ta	du

O sistema não acertou nem as sílabas tônicas nem a divisão correta das palavras em latim.

(2) O próximo verso ficou bastante estranho ao ouvido na escansão produzida pelo Aoidos:

MAT:9.2.6.3 de toda a iguaria isenta,

1	2	3	4	5	6	7	8
de	to	da	a	i	gua	ria	i
						sen	ta
di	'tɔ	da		'gwa	'ri	'zẽ	ta

O hiato entre *toda a* e *i* só é menos pior talvez que a elisão e subsequente crase entre *iguaria* e *isenta*. Uma escansão melhor seria:

1	2	3	4	5	6	7	8
de	to	da	i	gua	ri	a	i
						sen	ta

(3) O último verso de Gregório de Matos incorretamente escandido é este:

MAT:29.90.3.9 coma eu da fruta enfim,

1	2	3	4	5	6	7	8
co	ma	eu	da	frui	ta	en	fim
'ko	'mew	da	'fruj	ta	ẽ	'fĩ	

Aqui o hiato entre *fruta* e *enfim* é bastante questionável, assim como a elisão de *coma*, que, fundida à palavra seguinte, dá a impressão de que a forma verbal *comeu* está sendo dita. Uma escansão melhor evita esses problemas:

1	2	3	4	5	6	7	8
co	ma	eu	da	frui	ta	en	fim

(4) Este verso de Augusto dos Anjos foi escandido pelo sistema de forma quase aceitável:

ANJ.25.19.2 Adivinhando o frio que há nas lousas,

1 2 3 4 5 6 7 8 9 10 11  
 A-di-vi-nhan-do o fri-o que há nas lou-sas,  
 a di vi 'nã du 'fri:u 'kja naz 'low zas

Seria melhor se o sistema tivesse feito a sinérese em *frio* e o hiato entre *que* e *há*:

1 2 3 4 5 6 7 8 9 10 11  
 A-di-vi-nhan-do o fri-o que há nas lou-sas,

(5) Para este verso de Cláudio Manuel da Costa, o Aoidos também gerou uma boa escansão:

COS.11.4.3 Veja a Eulina; e então suspenda o pranto.

1 2 3 4 5 6 7 8 9 10 11  
 Ve-ja a Eu-li-na; e en- tão sus-pen-da o pran-to.  
 'vɛ ʒa ew 'li nɐ̃ tãw sus pẽ du 'prã tu

Contudo, seria preferível uma escansão em que fosse feita uma sinalefa com o início do nome *Eulina*:

1 2 3 4 5 6 7 8 9 10 11  
 Ve-ja a Eu-li-na; e en- tão sus-pen-da o pran-to.

(6) Do mesmo poeta é o verso seguinte, que o sistema escandiu de forma consistentemente errada:

COS.108.66.3 Que eu digo o meu, tu, lê o teu conceito,

1 2 3 4 5 6 7 8 9 10 11  
 Que eu di-go o meu, tu, lê o teu con-cei-to,  
 'kjew 'di gu u 'mew tu 'lew tew kō sej tu

O hiato entre *digo* e *o* não é tolerável. Melhor seria:

1 2 3 4 5 6 7 8 9 10 11  
 Que eu di-go o meu, tu, lê o teu con-cei-to,

(7) No verso que segue, o sistema fez a escolha errada de hiatos:

DUR.3.4.5 “— Diz-me, hóspede amigo, se professa

1 2 3 4 5 6 7 8 9 10 11  
 “— Diz-me, hós-pe-de a mi-go, se pro-fe-ssa  
 'diz 'mjɔs pɪ dɪ a 'mi gu sɪ pro 'fɛ sa

Para escandi-lo, é preciso fazer o hiato ou entre *Diz-me* e *hóspede* ou entre *hóspede* e *amigo*; a primeira escolha é melhor que a segunda:

1 2 3 4 5 6 7 8 9 10 11  
 “— Diz-me, hós-pe-de a mi-go, se pro-fe-ssa

(8) O verso seguinte faz parte de um poema polimétrico de Araújo Figueredo:

FIG:2.28.2.39 O cego, em êxtase:

1 2 3 4 5 6  
 O|ce-go, em|êx-ta-se:  
 u|sẽ gwẽ 'es-ta zi

O problema desta escansão não pode ser encontrado considerando-se o verso em isolamento. Ele está numa estrofe em que há alguns versos de 5 sílabas e outros tantos de 11 com acento na 5ª. A análise métrica do sistema não soube perceber isso. É melhor forçar a escansão um pouco para que ele esteja ritmicamente compatível com seus vizinhos:

1 2 3 4 5 6 7  
 O|ce-go, em|êx-ta-se:

(9) De Gonçalves de Magalhães é o verso seguinte, em que o leitor precisa decidir entre uma sinalefa e uma sinérese:

MAG.15.7.6 É tudo a experiência?

1 2 3 4 5 6 7  
 É|tu-do a|ex-pe-riên-cia?  
 'ɛ|tu|dwa|es|pe|rjẽ|sja

O sistema optou pela sinérese, mas com a sinalefa a escansão fica melhor:

1 2 3 4 5 6 7  
 É|tu-do a ex-pe-ri-ên-cia?

(10) Num verso da Divina Comédia o sistema se deparou com uma palavra em italiano:

PIN.12.37.3 Obizzio d'Este, ao qual (verdade eu digo)

1 2 3 4 5 6 7 8 9 10 11  
 O-biz- zio d'Es-te, ao qual |(ver-da-de eu|di-go)  
 o biz|ziw|'dɛs|tjaw|kwal|ver|da|djew|di|gu

O sistema errou na sílaba tônica de *Obizzio*, além de adotar uma pronúncia bastante contestável para ela.

(11) Caso semelhante é o do verso seguinte:

PIN.30.10.1 Uma alcançou Capocchio e, lhe cravando

1 2 3 4 5 6 7 8 9 10 11  
 U-ma al-can-çou|Ca-poc-chio e, lhe|cra-van-do  
 u|mal|kã|sow|ka|pok|ʃjwɪ|ɫi|kra|vã|du

Aqui o sistema foi capaz de fazer uma sístole em *Capocchio* e por isso acertou a sílaba tônica, mas desfigurou de tal maneira a pronúncia de *Capocchio* que foi considerado incorretamente escandido.

### 6.2.3. Discussão

**Regras.** Foram necessárias 159 regras para escandir os versos dos 10 *corpora*. A Tabela 6.5 mostra a quantidade de regras que cada *corpus* indivi-

Tabela 6.5.: Quantidade de regras aplicadas no processamento dos *corpora*. As *Utilizadas* são todas as que o sistema precisou para escandir o *corpus*; as *Exclusivas* são aquelas que apenas aquele *corpus* em particular empregou. As colunas %o mostram as contagens por mil versos. Os *Pontos* são a pontuação de naturalidade média das regras utilizadas pelo *corpus*.

<i>Corpus</i>	Versos	Utilizadas			Exclusivas		
		Total	%o	Pontos	Total	%o	Pontos
CAM	8.816	66	7,5	2,7	5	0,6	0,8
MAT	33.840	87	2,6	3,1	7	0,2	3,6
DUR	6.672	58	8,7	2,6			
COS	7.682	40	5,2	0,8	1	0,1	-1,0
GON	4.184	24	5,7	1,1			
MAG	5.637	47	8,3	2,1			
PIN	14.233	58	4,1	2,3	1	0,1	3,0
ANJ	6.594	92	14,0	2,8	22	3,3	4,0
FIG	8.233	77	9,4	3,5	8	1,0	6,0
SIL	11.547	79	6,8	4,1	12	1,0	7,2
	107.438	159	1,5	2,5	56	0,5	2,4

dual precisou, variando de 24 a 92. É razoável a expectativa de que, até um certo limite, quanto mais versos houver num *corpus*, maior será o número de regras; por esse motivo, a tabela mostra também o número de regras por 1.000 versos. A quantidade média de pontos das regras de cada *corpus* indica o quão extremas foram os fenômenos utilizados pelo poeta (pelo menos de acordo com as noções que o sistema tem acerca dos fenômenos): quanto mais alta a pontuação, mais extremas são as regras. A tabela mostra ainda as regras exclusivas a cada *corpus*, ou seja, regras que nenhum dos outros *corpora* utilizou.

É notável a discrepância que existe entre os *corpora*. Enquanto 87 regras foram suficientes para escandir os 33.840 versos de Gregório de Matos, os 6.594 de Augusto dos Anjos demandaram 92. Parece haver uma tendência de os poetas mais recentes exigirem um número maior de regras exclusivas, mas até mesmo a obra de Camões precisou de 5 regras exclusivas. O *corpus* de Cláudio Manuel da Costa destaca-se pela baixa média de pontuação das regras usadas, apenas 0,8, enquanto as regras de Delminda Silveira tiveram uma média de 4,1. O contraste é ainda maior quando as regras exclusivas são consideradas: o *corpus* COS precisou de



Tabela 6.6.: Número de aplicações de regras a cada *corpus* e de elocuições encontradas.

<i>Corpus</i>	Versos	Sílabas	Aplicações			Elocuições		
			Total	/Verso	/Sílaba	Total	/Verso	/Sílaba
CAM	8.816	88.160	25.924	2,94	0,29	233.767	26,52	2,65
MAT	33.840	246.716	65.770	1,94	0,27	410.344	12,13	1,66
DUR	6.672	66.720	19.953	2,99	0,30	199.908	29,96	3,00
COS	7.682	69.200	19.752	2,57	0,29	162.486	21,15	2,35
GON	4.184	41.708	11.369	2,72	0,27	102.813	24,57	2,47
MAG	5.637	49.039	15.066	2,67	0,31	154.398	27,39	3,15
PIN	14.233	142.342	43.385	3,05	0,30	448.347	31,50	3,15
ANJ	6.594	65.314	21.357	3,24	<b>0,33</b>	268.958	40,79	4,12
FIG	8.233	83.875	27.109	<b>3,29</b>	0,32	386.531	<b>46,95</b>	<b>4,61</b>
SIL	11.547	97.867	26.313	2,28	0,27	214.305	18,56	2,19
	107.438	950.941	275.998	2,57	0,29	2.581.857	24,03	2,72

uma regra exclusiva com pontuação -1, enquanto SIL necessitou 12 com pontuação média 7,2.

A Tabela 6.6 mostra quantas aplicações de regras e elocuições foram necessárias na escansão de cada *corpus*. Essas quantidades estão exibidas em números absolutos e também em relação ao número de versos e de sílabas. A quantidade de sílabas é relevante porque o número de aplicações de regras está relacionado ao tamanho do verso. Uma taxa maior de aplicação de regras por verso pode significar versos maiores, não necessariamente versos mais complexos. O *corpus* FIG, por exemplo, precisou em média de 3,29 regras por verso, enquanto o *corpus* ANJ precisou de 3,24; porém Araújo Figueredo escreveu versos maiores do que Augusto dos Anjos. Se considerarmos a taxa por sílaba, os versos do *corpus* ANJ parecem mais complexos, com 0,33 regras por sílaba, enquanto FIG precisou de 0,32. Essas taxas relativas provavelmente correspondem a características estilísticas.

O número médio de elocuições criadas pelo sistema para cada verso foi de 24,03. Novamente, há bastante variação entre os *corpora*. O *corpus* MAT destaca-se com apenas 12,13 elocuições por verso; do outro lado está FIG, com 46,95. A taxa por sílaba ajuda a reduzir diferenças aparentes entre os *corpora*: se considerarmos a taxa por verso, DUR precisou de 2,47 vezes mais elocuições que MAT; com a taxa por sílaba, a razão cai para 1,81. É provável que, devido à natureza combinatória do gerador de elocuições

Tabela 6.7.: Quantidade de *corpora* aos quais as regras foram aplicadas e o número de aplicações correspondente. A primeira linha, por exemplo, diz que 56 regras foram aplicadas a um único *corpus*.

<i>Corpora</i>	Regras	Aplicações		
		Total	/Regra	% Versos
1	56	86	1,5	0,8
2	24	149	6,2	1,4
3	20	170	8,5	1,6
4	4	28	7,0	0,3
5	9	134	14,9	1,2
6	4	72	18,0	0,7
7	6	312	52,0	2,9
8	7	423	60,4	3,9
9	9	1.744	193,8	16,2
10	20	272.880	13.644,0	2.539,9

do Aoidos, o número de elocuições não aumente de forma linear com o tamanho do verso, motivo pelo qual comparar o número de elocuições de diferentes *corpora* é menos útil do que o número de aplicações de regras.

A Tabela 6.5 informa que existem 56 regras exclusivas a um *corpus* no sistema, mas isso não significa que as demais regras são aplicadas a todos. A Tabela 6.7 contém justamente esta informação: cada linha mostra a quantidade de regras que foi aplicada ao número de *corpora* mostrado na primeira coluna. As três colunas finais contêm a quantidade de vezes que as regras foram de fato aplicadas, bem como a taxa de aplicações por cada regra e por mil versos dos *corpora*.

O que se percebe é que apenas 20 regras se aplicam a todos os *corpora*, mas elas são responsáveis pela maioria esmagadora das aplicações: 272.880 ou 98,87%. Pode-se dizer que existe um núcleo básico de regras, compartilhado entre todos os poetas e usado de forma maciça, e regras mais raras e específicas, usados por menos poetas e mais raramente. As regras pertencentes a esse núcleo básico encontram-se destacadas no Apêndice B com uma estrela. Também é notável que, conforme diminuem a quantidade de *corpora* aos quais as regras são aplicáveis, também diminui a taxa de aplicações por regra. Não se trata, portanto, de que alguns poetas utilizam muito algumas regras específicas suas. De fato, as 56 regras aplicáveis a somente um *corpora* foram aplicadas em média 1,5 vezes, ou seja, muitas delas foram aplicadas apenas uma vez. Não está claro se esta concentração de frequência em poucas regras reflete características do sis-

Tabela 6.8.: Resultado da validação cruzada: número de regras disponíveis e faltantes e quantidade de versos prejudicados se os *corpora* não excluídos servissem de treinamento para a avaliação do *corpus* excluído.

Excluído	Regras		Versos	
	Disponíveis	Faltantes	Prejudicados	%
CAM	154	5	7	0,8
MAT	152	7	8	0,2
DUR	<b>159</b>	<u>0</u>	<u>0</u>	
COS	158	1	1	0,1
GON	<b>159</b>	<u>0</u>	<u>0</u>	
MAG	<b>159</b>	<u>0</u>	<u>0</u>	
PIN	158	1	2	0,1
ANJ	<u>137</u>	<b>22</b>	<b>30</b>	<b>4,5</b>
FIG	151	8	12	1,5
SIL	147	12	25	2,2

tema ou da poesia.

Um questionamento relevante refere-se à capacidade do sistema de analisar *corpora* não incluídos no experimento, já que existem *corpora* que necessitam de regras específicas. A Tabela 6.8 mostra o resultado de uma espécie de validação cruzada: consideram-se somente as regras necessárias à escansão dos *corpora* não excluídos e avalia-se o impacto causado sobre o *corpus* excluído. Dito de outra forma, a tabela mostra quantos versos o sistema não conseguiria mais escandir sem as regras exclusivas do *corpus* excluído.

Com a Tabela 6.8, torna-se claro que a quantidade de trabalho para tratar um *corpus* novo depende das características do *corpus* em si. Enquanto três *corpora*, DUR, GON e MAG poderiam ser escandidos sem a adição de nenhuma regra, o *corpus* ANJ precisaria de 22 regras a mais. Se levarmos em consideração que a ampla maioria das aplicações das regras são feitas por algumas poucas e que, portanto, essas 22 regras a mais são de fato exclusivas e excepcionais, é provável que o número de versos prejudicados, 30, seja linearmente relacionado à quantidade de versos do *corpus*. É possível, portanto, que ao tratar um *corpus* muito maior do que o de Augusto dos Anjos mas com características estilísticas similares seja necessário adicionar uma boa quantidade de regras ao sistema.

A Tabela 6.9 mostra a quantidade de aplicações das regras, agrupadas por categorias, com as regras especiais excluídas. A quantidade mostrada é a taxa de aplicações em cada *corpus* por 10 mil sílabas. A utiliza-

Tabela 6.9.: Quantidade de aplicações das regras por 10 mil sílabas, agrupadas por categoria.

<i>Corpus</i>	Sinalefa	Elisão	Crase	Sinérese	Eclipse	Âferese	Síncope	Acento	Diérese	Redução	Sístole	Paragoge	Epêntese	Apócope	Diástole
CAM	670,1	253,5	142,4	18,4	21,7	1,9	0,7		1,5	0,5	0,1				
MAT	690,8	222,9	86,9	22,1	5,9	0,5	0,2	3,0	0,6	0,3	0,2			0,0	
DUR	946,2	323,9	143,6	56,8	4,3	3,0	0,6	0,1	1,0	0,1	0,3				
COS	857,1	296,5	112,6	32,4	2,5	1,2			0,3	0,1					
GON	574,7	190,6	65,9	9,8	1,2	0,2			0,5						
MAG	753,9	285,1	124,8	21,6	17,7	3,7	0,6	0,2	0,4	1,4			0,2		
PIN	920,2	323,6	129,6	25,2	1,8	2,8	0,1		0,6	0,6	0,4	0,1			
ANJ	870,9	336,7	171,2	182,3	29,1	2,4	0,6	0,9	0,5	0,2					
FIG	783,0	339,1	110,8	73,4	5,5	1,5	1,9	0,6	0,8	2,9	0,5	0,4		0,1	0,1
SIL	606,2	270,4	79,7	41,3	15,4	1,0	8,2	0,2	0,9	0,3	0,3	0,3	0,6		
	50,9	18,5	7,5	2,8	0,6	0,1	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0

ção da sílaba como denominador é uma tentativa de reduzir a influência do tamanho dos versos, permitindo comparações mais significativas entre os *corpora*. As colunas estão ordenadas pela taxa média de aplicação das regras de cada categoria.

Destaca-se na Tabela 6.9 o *corpus* DUR, com a maior quantidade de sinalefas por sílaba. O *corpus* GON parece conter os versos mais fáceis de serem escandidos: possui a menor taxa de aplicação por sílaba em todas as categorias, exceto diérese. É notável a grande quantidade de sinéreses praticadas por Augusto dos Anjos: ele faz 2,5 vezes mais que Araújo Figueredo, que está em segundo lugar. Este número alto de sinéreses talvez explique a quantidade relativamente baixa de sinalefas no *corpus* ANJ. Gregório de Matos utilizou as regras de acento muito mais do que todos os outros; isto se deve à grande quantidade de versos terminados em *que*, *porque*, etc., como no seguinte:

MAT:29.107.1.3 que me deram junto ao que

1 2 3 4 5 6 7 8

que|me|de-|ram|jun-|to ao|que|

kɪ mɪ 'dɛ rãw 'zũ twaw 'ke

É importante notar que, para vários dos fenômenos da Tabela 6.9, a taxa exibida não reflete necessariamente a quantidade real presente nos versos. Ela reflete a quantidade de fenômenos *não marcados graficamente*

Tabela 6.10.: Pontuação média das aplicações de regra, segundo a categoria. Somente categorias com pelo menos 100 aplicações estão incluídas.

<i>Corpus</i>	Sinalefa	Elisão	Crase	Sinérese	Eclipse	Afêrese	Síncope
CAM	-2,43	-2,09	-1,88	<b>2,15</b>	1,25	<u>4,00</u>	6,67
MAT	-2,55	-2,13	-1,92	0,58	1,64	<u>4,92</u>	<u>5,50</u>
DUR	-2,35	<b>-2,06</b>	-1,94	1,00	1,24	<u>4,00</u>	7,00
COS	-2,51	-2,15	-1,96	0,54	<u>1,00</u>	<u>4,00</u>	
GON	<u>-2,63</u>	-2,07	-1,97	1,37	<u>1,00</u>	<u>4,00</u>	
MAG	-2,45	<u>-2,17</u>	-1,95	1,26	1,03	4,61	6,00
PIN	-2,43	<u>-2,13</u>	-1,91	1,32	1,72	4,28	6,00
ANJ	<b>-2,15</b>	-2,14	<b>-1,80</b>	0,72	1,21	<b>5,38</b>	7,00
FIG	-2,41	-2,14	-1,86	1,22	<b>3,67</b>	<u>4,00</u>	<b>7,88</b>
SIL	-2,44	-2,08	<u>-1,98</u>	<u>0,52</u>	3,03	5,10	6,81
	-2,44	-2,12	-1,92	1,07	1,68	4,43	5,29

*no texto.* Existe um motivo para o grandíssimo número de síncope re-realizadas por Delminda Silveira: elas não estão marcadas no texto. No seguinte verso, por exemplo, o sistema faz uma síncope em *esperanças*:

SIL:5.124.3.1 Eu também te saúdo, ano d'esperanças...

1 2 3 4 5 6 7 8 9 10 11  
 Eu | tam- | bém | te | sa- | ú- | do, a- | no | d'es- | peran- | ças...  
 'ew | tã | 'bê | tr | sa | 'u | dwã | nu | des | 'prã | sas

As síncope acontecem também com bastante frequência no *corpus* PIN, mas a edição as costuma marcar. A contagem de 0,1 síncope por 10 mil sílabas não inclui, por exemplo, o seguinte verso, onde há síncope na mesma palavra *esperança*:

PIN.5.15.2 De pausa não, de menos dor a esp'rança

1 2 3 4 5 6 7 8 9 10 11  
 De | pau- | sa | não, | de | me- | nos | do- | r a es- | p'ran- | ça  
 dɪ | 'paw | za | 'nãw | dɪ | 'me | nuz | 'dɔ | raɣs | 'prã | sa

A pontuação das regras do sistema varia mesmo entre as que pertencem a uma mesma categoria. Desprezando-se a quantidade de aplicações das regras e considerando-se a pontuação em cada aplicação das regras, a Tabela 6.10 mostra o quão extremas são as regras utilizadas pelos diferentes poetas. As colunas da tabela seguem a mesma ordem da tabela anterior, segundo o número médio de aplicações por sílaba. As sinalefas de Tomás Antônio Gonzaga, por exemplo, são as mais bem-vindas entre todos

Tabela 6.11.: Quantidade de versos de acordo com o número de sílabas poéticas, em valores absolutos.

	1	2	3	4	5	6	7	8	9	10	11	12	13+	O.	Total
CAM										8.816					8.816
MAT	15	50	220	243	319	386	28.318	21	107	4.059	100	1		1	33.840
DUR										6.672					6.672
COS		2		37	6	1.663	231		7	5.736					7.682
GON										4.172				12	4.184
MAG				321	128	782	540	3	46	3.782	35				5.637
PIN										14.225	4	4			14.233
ANJ				3		77	276		21	5.834	9	358		16	6.594
FIG	1	43	23	80	71	227	1.027	106	144	2.759	290	3.444	18		8.233
SIL		1	123	428	405	486	3.920	69	396	4.492	317	894	12	4	11.547
	16	96	366	1.112	929	3.621	34.312	199	721	60.547	755	4.701	30	33	107.438

Tabela 6.12.: Quantidade de versos de acordo com o número de sílabas poéticas, em percentual.

	1	2	3	4	5	6	7	8	9	10	11	12	13+	O.	Total
CAM										100,0					100
MAT	0,0	0,1	0,7	0,7	0,9	1,1	83,7	0,1	0,3	12,0	0,3	0,0		0,0	100
DUR										100,0					100
COS		0,0		0,5	0,1	21,6	3,0		0,1	74,7					100
GON										99,7				0,3	100
MAG				5,7	2,3	13,9	9,6	0,1	0,8	67,1	0,6				100
PIN										99,9	0,0	0,0			100
ANJ				0,0		1,2	4,2		0,3	88,5	0,1	5,4		0,2	100
FIG	0,0	0,5	0,3	1,0	0,9	2,8	12,5	1,3	1,7	33,5	3,5	41,8	0,2		100
SIL		0,0	1,1	3,7	3,5	4,2	33,9	0,6	3,4	38,9	2,7	7,7	0,1	0,0	100
	0,0	0,1	0,3	1,0	0,9	3,4	31,9	0,2	0,7	56,4	0,7	4,4	0,0	0,0	100

os poetas, com pontuação média de 2,63; Augusto dos Anjos fica no outro extremo, com pontuação média para as sinalefas de 2,15. O fato de que da esquerda para a direita a pontuação média vai aumentando evidencia que, em geral, regras mais raramente utilizadas possuem uma pontuação maior.

**Versos.** As Tabelas 6.11e e 6.12 mostram a contagem de versos de acordo com o número de sílabas poéticas para todos os *corpora*, respectivamente

em valores absolutos e relativos. A coluna 13+ agrupa todos os versos que contenham 13 ou mais sílabas. A coluna O. contém os outros casos: versos para os quais o sistema não produziu nenhuma sílaba. Fica imediatamente claro quais são as obras compostas com um único tamanho de verso: os *corpora* CAM, DUR, GON e PIN contêm exclusivamente decassílabos. Entre os demais *corpora*, todos apresentam uma certa variedade, com MAT e FIG cobrindo boa parte do espectro.

A Tabela 6.11 contém, contudo, contradições evidentes. Existem grandes poemas decassílabos com versos de 12 ou 13 sílabas; há *corpora* como FIG e SIL com versos com 13 ou mais sílabas; existe um único dodecassílabo na obra de Gregório de Matos. Estas contradições se devem a dois fatores principais: a) a presença de 130 versos inescandíveis nos *corpora*, que foram considerados como tais justamente por não se adaptarem à métrica do poema em que estão contidos; b) erros na escansão do sistema, que se devem especialmente a uma análise métrica incorreta. Os números relativos exibidos na Tabela 6.12, por outro lado, indicam que as contradições, embora não desprezíveis, são poucas diante do número total de versos analisados.

### 6.3. Experimento secundário de validação

Este experimento, em vez de utilizar versos variados de uma quantidade proporcionalmente pequena de vários *corpora*, visa a avaliar versos de apenas dois tipos contidos num mesmo tipo de poema de um mesmo poeta: são os decassílabos e dodecassílabos que Augusto dos Anjos utilizou em seus sonetos. Foram escolhidos para serem avaliados 111 sonetos entre os compostos pelo poeta, que contêm um total de 1.554 versos ou 23,6% de sua produção poética completa. Apenas 2 sonetos destes usam dodecassílabos, os 109 restantes estando escritos em decassílabos.

O experimento havia sido realizado com uma versão anterior do sistema (131) e os versos cuja escansão se alterou da versão originalmente utilizada para a atual foram manualmente reavaliados.

**Crítérios.** Cada soneto teve seus versos avaliados de forma independente por dois especialistas. Foi-lhes solicitado que respondessem, para cada verso, à pergunta: *Esta é uma escansão aceitável?* Como o critério de aceitabilidade é subjetivo, uma dada escansão foi considerada inaceitável se pelo menos um dos dois especialistas a classificaram como tal. O critério de avaliação difere daquele do experimento principal porque eles não fo-

ram executados simultaneamente; o experimento principal, realizado mais tarde, utiliza um critério de avaliação menos subjetivo. Os avaliadores tiveram acesso a uma interface especificamente desenvolvida para o experimento, onde eles podiam visualizar a transcrição fonética e as vogais marcadas como tônicas pelo sistema. Os avaliadores podiam deixar comentários a cada verso sempre que julgassem necessário, independentemente da classificação atribuída a ele.

**Resultados.** Dos 1.554 versos, 11 versos tiveram sua escansão rejeitada como inaceitável por pelo menos um especialista, o que equivale a uma taxa de acerto de 99,3%, um pouco acima daquela encontrada no experimento principal, de 98,5%. Entretanto, enquanto o presente experimento tem uma boa cobertura da obra do poeta, ainda que apenas dos sonetos, o outro avaliou apenas 65 versos, 1% da produção do poeta. Cabe ressaltar que os critérios de avaliação não foram os mesmos nos dois experimentos e que, desta sorte, a comparação das taxas de acerto deve ser feita de forma cuidadosa. Os versos considerados inaceitáveis, por serem poucos, são discutidos individualmente a seguir.

(1) No seguinte verso, o sistema preferiu deixar a forma verbal *é* na sexta sílaba, enquanto os especialistas afirmam que é preferível que a sílaba tônica de *Humana* ocupe aquela posição:

ANJ.3.4.1 A Consciência Humana é este morcego!

1	2	3	4	5	6	7	8	9	10	11
A	Con	sciên	cia Hu	ma	na é	es	te	mor	ce	go!
a	kõ	'sjẽ	sju	'mã	'nẽ	'es	ti	mor	'sẽ	gu

A escansão corrigida seria:

1	2	3	4	5	6	7	8	9	10	11	
A	Con	sci	ên	cia Hu	ma	na é	es	te	mor	ce	go!

(2) O sistema encontrou as 10 sílabas do seguinte verso deixando átona a palavra *Um*, o que produziu uma escansão deveras estranha:

ANJ.19.2.3 É a morte, é esse danado número Um

1	2	3	4	5	6	7	8	9	10	11	12	
É	a	mor	te,	é	e	sse	da	na	do	nú	me	ro Um
'ε	a	'mɔɾ	'tʃɛ	'ɛ	sɪ	da	'na	du	'nu	mi	rwũ	

(3) Já neste verso o sistema preferiu fazer uma eclipse em vez de uma sinalefa mais complexa na primeira sílaba do verso, quando o contrário seria mais adequado:

ANJ.19.4.2 Que o homem universal de amanhã vença



1 2 3 4 5 6 7 8 9 10 11  
 Que o h-o-mem u-ni-ver-sal de a-ma-nhã ven-ça  
 kju 'o mju ni ver'sal dja ma 'nã 'vẽ sa

A escansão preferível seria:

1 2 3 4 5 6 7 8 9 10 11  
 Que o ho-mem u-ni-ver-sal de a-ma-nhã ven-ça

(4) De exatamente o mesmo problema sofre o próximo verso:

ANJ.35.1.3 O que o homem ama e o que o homem abomina,

1 2 3 4 5 6 7 8 9 10 11  
 O que o h-o-mem a-ma e o que o ho-mem a-bo-mi-na,  
 u kju 'o 'mjã mju 'kjwo mẽ a bo 'mi na

(5) Aqui o sistema escolheu a sinérese em vez da sinalefa, quando o contrário seria preferível:

ANJ.120.2.4 Trovador torturado e angustioso.

1 2 3 4 5 6 7 8 9 10 11  
 Tro-va-dor tor-tu-ra-do e an-gus-tio-so.  
 tro va 'dɔr tor tu 'ra du jã gus 'tjo zu

A escansão corrigida seria:

1 2 3 4 5 6 7 8 9 10 11  
 Tro-va-dor tor-tu-ra-do e an-gus-ti-o-so.

(6) Neste verso a escansão do sistema deixou na 5ª sílaba uma palavra importante do verso, *azul*, e na 6ª uma preposição, *ante*, o que forçaria uma escansão heroica:

ANJ.132.2.4 Vencendo o azul que ante si s'erguera.

1 2 3 4 5 6 7 8 9 10 11  
 Ven-cen-do o a-zul que an-te si s'er-gue-ra.  
 vẽ 'sẽ du wa 'zul 'kjã 'tri si ser 'ge ra

Mais apropriada seria uma escansão sáfica em que *azul* e *si* recebem os ictos:

1 2 3 4 5 6 7 8 9 10 11  
 Ven-cen-do o a-zul que an-te si s'er-gue-ra.

(7) Novamente o sistema prefere uma sinérese a uma sinalefa, quando o contrário seria mais aceitável:

ANJ.132.3.3 Aureoreando a humana consciência.

1 2 3 4 5 6 7 8 9 10 11  
 Au-ro-rean-do a h-u-ma-na con-sci-ên-cia.  
 aw ro 'rjã dwa u 'mã na kô si 'ẽ sja

Aqui a palavra *Aureoreando* deve ser pronunciada em 5 sílabas, deixando seu acento na 4ª posição:

1 2 3 4 5 6 7 8 9 10 11  
 Au-ro-re-an-do a hu-ma-na con-sci-ên-cia.

(8) Neste verso, um dos especialistas julgou ser mais aceitável uma escansão que leve em conta a pontuação:

ANJ.146.1.1 Tarde. Um arroio canta pela umbrosa

1	2	3	4	5	6	7	8	9	10	11		
Tar-	de.	Um	a-	rroi-	o	can-	ta	pe-	la	um-	bro-	sa
'tar		djũ		a	'boj	u	'kã	ta	'pe	lũ	'brɔ	za

Ao privilegiar o ponto após *Tarde*, o artigo *Um* sofreria eclipse e a consequente sinalefa com a palavra seguinte:

1	2	3	4	5	6	7	8	9	10	11		
Tar-	de.	Um	a-	rroi-	o	can-	ta	pe-	la	um-	bro-	sa

(9) O sistema neste verso não foi capaz de fazer uma sinérese em *continue*, o que o levou a fazer uma outra sinérese, bem mais forçada, na palavra *lua*, desfigurando o ritmo deste heroico:

ANJ.210.4.3 A lua continue sempre a nascer!

1	2	3	4	5	6	7	8	9	10	11
A	lua	con-	ti-	nu-	e	sem-	pre	a	na-	scer!
a	'lwa	kô	ti	'nu	ɪ	'sẽ	prja	na	'ser	

A escansão preferível seria:

1	2	3	4	5	6	7	8	9	10	11
A	lu-	a	con-	ti-	nue	sem-	pre	a	na-	scer!

(10) Um dos especialistas indicou que, no verso seguinte, seria melhor que o acento da palavra *clara* ocupasse a prestigiosa 6ª posição:

ANJ.243.1.1 Vibram na sua voz clara de criança,

1	2	3	4	5	6	7	8	9	10	11
Vi-	bram	na	su-	a	voz	cla-	ra	de	crian-	ça,
'vi	brãw	na	'su	a	'vos	'kla	ra	dɪ	'krjã	sa

Para obter a escansão sugerida, faz-se necessária uma sinérese em *sua* e uma diérese em *criança*:

1	2	3	4	5	6	7	8	9	10	11
Vi-	bram	na	sua	voz	cla-	ra	de	cri-	an-	ça,

(11) Já neste verso, seria mais adequado uma pronúncia com diérese e subsequente sinalefa em *tábua*, evitando o hiato entre a vogal final desta palavra e o artigo *o*:

ANJ.280.2.1 Para a tábua o mundo inteiro manda

1	2	3	4	5	6	7	8	9	10	11		
Pa-	ra	a	tá-	bua	o	mun-	do	in-	tei-	ro	man-	da
'pa	ra	'ta	bwa	u	'mũ	dwĩ	'tej	ru	'mã	da		

A escansão desejável seria:

1	2	3	4	5	6	7	8	9	10	11			
Pa-	ra	a	tá-	bu-	a	o	mun-	do	in-	tei-	ro	man-	da

#### 6.4. Caso de uso: os versos de Gregório de Matos

Este caso de uso relata a experiência de aplicação do sistema ao *corpus* da poesia completa de Gregório de Matos e a comparação dos resultados obtidos com aqueles produzidos manualmente por Chociay (74). Este é um livro que, além de conter estatísticas sobre os versos do poeta, contém ricas discussões de ordem versificatória. Chociay é autor também do influente manual *Teoria do Verso* (75).

Chociay utilizou a 2ª edição de James Amado em sua pesquisa, enquanto a edição de referência adotada nesta tese é a 3ª. Não parece haver grandes diferenças entre as duas edições, a julgar pelas menções de páginas feitas no livro de Chociay, que coincidem com a 3ª edição. O livro também deixa claro que, apesar de a de James Amado ser a edição básica da pesquisa, também foi consultada a da Academia Brasileira de Letras, o que ajudou o autor a estabelecer a versão correta de vários versos e também a suplementar poemas com versos faltantes.

Cabe ressaltar que existem questões de autoria não resolvidas no que se refere à obra de Gregório de Matos. Hansen (104) afirma que não apenas o poeta não publicou nenhum poema em vida como a própria concepção de autoria à época em que ele viveu difere da hodierna, de modo que a autoria dos poemas comumente atribuídos a Gregório não é firmemente estabelecida. Neste caso de uso, quando se fala de Gregório de Matos, fala-se do Gregório de Matos *da edição de James Amado*. O importante aqui é que tanto Chociay quanto o Aoidos conversem com esse mesmo Gregório de Matos.

**Os números de Chociay.** De interesse à presente tese são os números mencionados por Chociay ao longo do livro. Não é o objetivo deste experimento buscar e conferir todos os números lá apontados, especialmente porque eles não se encontram concentrados em alguma parte específica: estão espalhados ao longo dele. Os seguintes números mencionados nas respectivas páginas por Chociay serão comentados:

- p. 51: Gregório de Matos escreveu em espanhol 21 poemas, compostos de 78 estrofes e 584 versos, sendo 472 heptassílabos e 112 decassílabos;
- p. 67: existem 1.837 décimas espinelas em heptassílabos e 1.672 quadras heptassilábicas assonantadas;
- p. 67: o poeta escreveu um total de 196 sonetos;

Tabela 6.13.: Tabela de versos e frequências na obra de Gregório de Matos, retirada de Chociay (74).

Versos	Número de vezes	%
Dissílabo	15	0,04
Redondilho quebrado	251	0,76
Quebrado de <i>ovillejo</i>	66	0,20
Tetrassílabo	95	0,29
Pentassílabo (redondilho menor)	317	0,96
Hexassílabo	168	0,51
Heróico quebrado	123	0,37
Verso de seguidilha	52	0,16
Heptassílabo (redondilho maior)	28.028	84,61
Eneassílabo	96	0,30
Decassílabo italiano	3.798	11,46
Decassílabo 5-10	16	0,05
Decassílabo 4-7-10	10	0,03
Hendecassílabo	88	0,26
Totais gerais	33.123	100

- p. 52: os tipos de versos e sua frequência absoluta e relativa constam de uma tabela, reproduzida aqui na Tabela 6.13.

**Concordâncias e discordâncias.** Durante a revisão pontual dos *corpora*, poemas em língua estrangeira foram marcados no *corpus* como tal. Portanto, obter a contagem de poemas em espanhol é tarefa fácil; existem de fato 21 poemas escritos naquele idioma, como Chociay menciona. A contagem de estrofes no *corpus* desta tese foi de 90, 12 a mais do que o número listado por Chociay; no entanto, a separação dos versos em estrofes, tanto no *corpus* MAT quanto nos demais, mostrou-se bastante irregular na comparação entre as edições digitais e as de referência. Além disso, a própria edição de James Amado não costuma separar em estrofes todos os tipos de poema, mesmo aqueles que claramente contém quadras ou décimas. Comparar as contagens de estrofes obtidas por Chociay e presentes no *corpus* da tese, portanto, não é produtivo.

Apesar de haver no *corpus* MAT 21 poemas em espanhol, em concordância com Chociay, o mesmo não se pode dizer do número total de versos: Chociay afirma ter escrito o poeta 584 versos em poemas em espanhol, enquanto uma análise automática do *corpus* MAT revela 555 versos, 29 a menos que Chociay; mas se forem incluídos na contagem dos versos

em espanhol aqueles dos motes — que para todos os outros propósitos estão sendo ignorados nesta tese —, então uma contagem automática aponta 597 versos, 13 a mais que Chociay. É possível que o livro esteja, excepcionalmente, contando os referidos motes. Para harmonizar a contagem automática é aquela do livro, começa-se por considerado o fato de que o livro divide o número de 584 versos entre 472 heptassílabos e 112 decassílabos. Um exame manual dos 21 poemas em espanhol em MAT mostra, entretanto, que existem 9 sonetos decassilábicos escritos naquele idioma, como mostra a Tabela 6.14. Como cada soneto contém 14 versos, existem  $9 \times 14 = 126$  versos decassilábicos em espanhol, 14 a mais do que os 112 contados por Chociay, que aparentemente, portanto, pulou um soneto. Ainda assim, se somarmos os 14 versos do soneto ignorado na contagem de Chociay, o total de versos dele seria 598, o que deixa ainda uma diferença de um verso com relação à contagem obtida no *corpus* MAT — incluídos os motes —, de 597 versos. Uma busca no *corpus* MAT mostra que existe um poema heptassilábico, MAT:10.25, com uma estrutura idêntica a vários outros escritos por Gregório de Matos: um mote seguido de 4 décimas. Entretanto, falta um verso na terceira décima: uma análise das rimas revela que é o 6º ou 7º verso. A ausência deste verso não é um erro da edição digital; às pp. 422-3 da edição de James Amado pode-se constatar sua falta. Chociay, provavelmente por ter consultado a edição da Academia Brasileira de Letras, teve acesso ao verso em questão e o contou, o que explica o verso a mais em sua contagem. Não está claro, contudo, por que o número de poemas em espanhol encontrados por Chociay coincide com a contagem automático do *corpus* MAT, visto que há evidências de que ele tenha pulado um soneto.

O número de décimas espinelas e de quadras assonantadas informado por Chociay é de difícil verificação, porque requer a análise das rimas, o que não é feito pelo Aoidos. Uma décima espinela tem esquema rímico ABBAACCCDDC, mas Gregório de Matos escreveu décimas que não seguem este padrão. As quadras assonantadas não possuem rimas, mas o poeta compôs quadras com rimas. Os números apontados por Chociay, portanto, não foram verificados.

Já o número de sonetos pode ser verificado com facilidade devido à sua estrutura estrófica padronizada. Enquanto Chociay conta 196 deles, uma contagem automática produz um número mais alto: 217 sonetos, ou 21 a mais. Trata-se de uma diferença grande, de 294 versos. Para se ter certeza de que a quantidade de sonetos é de fato pelo menos 217, uma conferência manual foi feita na edição de James Amado: cada soneto acu-

Tabela 6.14.: Sonetos decassilábicos escritos em espanhol no *corpus* MAT, com referências às páginas em que aparecem na 2ª edição de James Amado.

Código	Página	Título
MAT:2.12	76	AO MISTERIOSO EPÍLOGO DOS INSTRUMENTOS...
MAT:2.13	76	RENDE-SE A PESSOA DE BERNARDO VIEYRA...
MAT:4.14	158	AO MESMO ASSUMPTO.
MAT:5.36	242	A FR. PASCOAL QUE SENDO ABBADE...
MAT:13.16	521	SEGUNDA LISONJA EM QUE EXCEDE SUA ESPOSA...
MAT:13.17	521	PRIMEIRO ARRUFOS DE SUA ESPOSA POR CAUSAS, ...
MAT:13.26	529	LISONJEIA FINALMENTE A CONVALESCENÇA...
MAT:26.22	767	COMPARA SUAS PENAS COM AS ESTRELLAS...
MAT:37.2	1180	DESCREVE O QUE REALMENTE SE PASSA...

sado pela análise automática do *corpus* foi encontrado na edição impressa. O resultado está na Tabela 6.15. Gregório de Matos escreveu, de fato, pelo menos 217 sonetos decassilábicos — deixando aberta a possibilidade de que haja sonetos na edição impressa que não aparecem na digital. Entre eles destacam-se 4 de cabo roto (mencionados por Chociay à p. 88), um com rimas dobradas (idem à p. 90) e um com disposição gráfica peculiar (idem à p. 123). Está evidente que os 217 sonetos contados automaticamente são legítimos; mesmo subtraindo a este número os 9 em espanhol, a diferença continuaria grande: 12 sonetos. Ao que tudo indica, a contagem de Chociay pulou os 21 sonetos.

**Heurísticas.** Os números da Tabela 6.13 não são muito fáceis de serem reproduzidos, porque os versos não são classificados somente segundo o seu tamanho. Os quebrados de *ovillejo*, por exemplo, têm um número variável de sílabas; os versos de seis sílabas são contados por Chociay como hexassílabos, heroicos quebrados ou versos de seguidilha, conforme o contexto em que aparecem. Algumas heurísticas de análise foram desenvolvidas para tentar reproduzir os critérios utilizados por Chociay na confecção de sua tabela.

Foram classificados como redondilhos quebrados aqueles versos em que o Aoidos encontrou 3 ou 4 sílabas poéticas e que se encontram em estrofes com pelo menos 2 versos heptassilábicos. Na seguinte estrofe, por exemplo, o último verso foi classificado como redondilho quebrado:

Tabela 6.15.: Páginas em que se encontram os sonetos de Gregório de Matos na edição de James Amado.

33	33	66	67	67	68	68	69	69	76
76	77	77	78	80	81	81	84	85	88
123	123	124	125	125	126	129	133	137	137
138	146	157	157	158	158	159	162	163	167
175	175	177	186	187	188	188	189	197	198
198	199	199	200	204	206	210	215	228	229
239	242	242	243	275	318	319	319	320	320
321	322	333	347	370	402	403	404	406	406
407	407	410	411	411	412	415	415	420	424
425	426	426	431	432	432	447	507	507	510
511	511	514	517	518	519	520	521	521	527
529	537	543	543	544	544	546	555	557	558
560	568	591	627	627	634	639	639	640	641
641	648	649	649	650	650	657	662	662	678
713	719	725	747	747	748	749	749	751	752
760	761	766	767	768	792	802	823	838	840
843	851	852	860	860	868	868	877	878	879
879	880	880	883	888	899	900	901	910	924
930	987	1045	1056	1071	1075	1112	1113	1113	1117
1120	1120	1129	1130	1137	1138	1163	1170	1180	1191
1192	1195	1202	1202	1203	1203	1204	1204	1205	1206
1206	1209	1211	1215	1216	1216	1222			

	1	2	3	4	5	6	7	8
MAT:6.13.1.1	Pois	me	dei	xais	pe-	lo	jo-	go,
MAT:6.13.1.2	li-	cen-	ça	me ha-	veis	de	dar	
MAT:6.13.1.3	pa-	ra	vos	sa-	ti-	ri-	zar	
MAT:6.13.1.4	co-	mo a-	mi-	go.				

Os poemas com *ovillejos* foram detectados da seguinte forma: considerou-se que poemas com um número  $n$  de estrofes,  $n$  sendo múltiplo de dois, em que as estrofes ímpares contém 6 versos e as pares contém 4, contam com  $n/2$  *ovillejos*. Assim, os versos pares das estrofes ímpares foram classificados como quebrados de *ovillejos*, independentemente do seu número de sílabas. Este é, por exemplo, o 2º *ovillejo* do poema MAT:30.19:

MAT:30.19.3.1 Que são da testa as carcomas?

MAT:30.19.3.2 Gomas

MAT:30.19.3.3 Ela diz, que são vertiges

MAT:30.19.3.4 Impiges

MAT:30.19.3.5 E lá dentro das alcobas?

MAT:30.19.3.6 Bobas,

MAT:30.19.4.1 Bem merece um par de sobas,

MAT:30.19.4.2 pois com quantos se pespega,

MAT:30.19.4.3 cada qual deles lhe pega

MAT:30.19.4.4 Gomas, Impiges, e Bobas.

Nesse *ovillejo*, os versos MAT:30.19.3.2, MAT:30.19.3.4 e MAT:30.19.3.6 foram classificados como quebrados de *ovillejo*.

Os versos de seis sílabas podem, na tabela de Chociay, serem classificados em três categorias distintas: hexassílabo, heroico quebrado ou verso de seguidilha. A categoria de hexassílabos refere-se a versos de seis sílabas em poemas isométricos — à possível exceção dos estrilhos, em que outros metros podem eventualmente ser encontrados. O sistema classificou como hexassílabos os versos de seis sílabas que se encontram em estrofes em que 75% ou mais dos versos contém seis sílabas. Os versos de seguidilha incluem versos de 4 e 6 sílabas; qualquer verso com um desses dois comprimentos foi classificado como de seguidilha quando localizados numa estrofe que contém ao menos 2 de quatro sílabas e 2 de seis sílabas. Os versos de seis sílabas que não atenderam os critérios de hexassílabos e de seguidilha foram classificados como quebrados de heroico.

Os decassílabos foram divididos em Chociay em três categorias: os italianos, os com esquema rítmico 5-10 e os com esquema 4-7-10. As heurísticas adotadas foram sequencialmente as seguintes: os decassílabos em que o sistema encontrou a 6ª sílaba ou então a 4ª e a 8ª sílabas acentuadas foram sumariamente classificados como italianos; os versos em que a 5ª sílaba estava acentuada foram classificados como 5-10; aqueles em que a 4ª e a 7ª estavam foram categorizados como 4-7-10; versos que não se encaixaram em nenhum desses padrões foram classificados como italianos.

**Comparação entre Chociay e resultados automáticos.** Os versos que não foram tratados por nenhuma heurística foram simplesmente classificados de acordo com seu número de sílabas poéticas. O resultado final pode ser avaliado na Tabela 6.16. A primeira constatação é que o sistema encontrou uma quantidade significativa de versos a mais do que Chociay: 1.272. Este número se deve sobretudo à contagem maior de heptassíla-



Tabela 6.16.: Resultados obtidos pelo Aoidos em comparação com a análise manual de Chociay. A porcentagem é a proporção da diferença com relação à contagem do sistema.

Classe	Aoidos	Chociay	Diferença	%
Dissílabo	15	15	0	0,0
Trissílabo	3		3	100,0
Redondilho quebrado	256	251	5	2,0
Quebrado de <i>ovillejo</i>	75	66	9	12,0
Tetrassílabo	155	95	60	38,7
Pentassílabo (redondilho menor)	319	317	2	0,6
Hexassílabo	168	168	0	0,0
Heróico quebrado	194	123	71	36,6
Verso de seguidilha	48	52	-4	-8,3
Heptassílabo (redondilho maior)	28.747	28.028	719	2,5
Octossílabo	21		21	100,0
Eneassílabo	107	96	11	10,3
Decassílabo italiano	4.141	3.798	343	8,3
Decassílabo 5-10	30	16	14	46,7
Decassílabo 4-7-10	14	10	4	28,6
Hendecassílabo	100	88	12	12,0
Dodecassílabo	1		1	100,0
Erro	1		1	100,0
	34.395	33.123	1.272	3,7

bos e decassílabos. A diferença de decassílabos pode ser bastante reduzida se assumirmos que Chociay de fato pulou 21 sonetos, ou seja, 294 versos. É difícil estabelecer se a contagem manual ignorou os heptassílabos, pois são muitos; mas entre os versos menos usados, pode-se verificar se houve alguma omissão.

A análise automática encontrou 75 quebrados de *ovillejo*, enquanto Chociay contou 66. Às pp. 99-103 de seu livro, ele afirma que quatro poemas foram escritos no formato de *ovillejos*, mencionando inclusive as páginas em que figuram na edição de James Amado: 55-6, 56-8, 830 e 1051-2. Uma contagem manual dos quebrados de *ovillejo* destes poemas mostra que eles contêm, respectivamente, 18, 27, 9 e 21 versos desse tipo, totalizando 75, o valor encontrado pelo sistema. A diferença entre este número e o mencionado por Chociay é 9, precisamente o número de quebrados no poema da p. 830, o que leva a crer que este poema foi pulado.

Os tetrassílabos encontrados pela análise automática foram 155, enquanto Chociay contou 95. Porém apenas dois poemas, MAT:4.35 e

MAT:29.23, contém respectivamente 84 e 57 tetrassílabos, o que pode ser conferido às pp. 183-186 e 846-849 da edição de James Amado. Esses dois poemas somam 141 versos, que são 46 a mais do que contou Chociay. Se assumirmos que foi pulado o poema de 57 tetrassílabos, a diferença entre a análise automática e a contagem de Chociay cai para apenas 3 versos, muito menor. Independentemente da razão para este resíduo, resta o fato de que existem no mínimo 141 tetrassílabos na obra de Gregório de Matos.

Se está demonstrado que Chociay deixou de contar ao menos 294 versos decassílabos oriundos de sonetos, 9 quebrados de *ovillejo* e 46 tetrassílabos, é possível que ele tenha ignorado também certa quantidade de heptassílabos, ainda que seja difícil apontar a origem dos versos ignorados. Com efeito, apesar do alto número absoluto da diferença de heptassílabos, 719, este número corresponde a apenas 2,5% da contagem automática, enquanto a mesma proporção de decassílabos italianos é 8,3%. Se os 294 versos dos sonetos forem adicionados à contagem de Chociay, a diferença de decassílabos italianos passa de 343 para 49 versos, trazendo a proporção deles de 8,3% para 1,2%; mesmo se considerarmos que são na verdade decassílabos italianos os eneassílabos encontrados a mais pelo sistema, bem como os decassílabos de outros tipos e os hendecassílabos, o excedente seria de 90 versos, ou 2,2%. Se uma diferença de 8,3% pode ser explicada por versos pulados, então é plausível que a diferença de 2,5% também possa ser explicada dessa forma.

Cabem ainda algumas observações a respeito das diferenças de contagem. A categoria de trissílabos, inexistente, contém 2 versos redondilhos quebrados que a heurística não conseguiu classificar como tal e um verso que parece de fato ser trissílabo, no seguinte estribilho:

	1	2	3	4	5	6	7
MAT:26.10.16.1	Oh	nun-	ca	se-	me-	lhan-	te
MAT:26.10.16.2	fo-	ra	eu	des-	ta	ro-	ca,
MAT:26.10.16.3	oh	nun-	ca	fo-	ram	tan-	tos
MAT:26.10.16.4	nem	tão	for-	tes	meus	ma-	les
MAT:26.10.16.5	co-	mo a-	s on-	das.			

A única categoria para a qual Chociay encontrou mais versos que o sistema foi a de versos de seguidilha, o que se deve ao fato de que as heurísticas não identificaram como versos de seguidilha os componentes desta estrofe:

	1	2	3	4	5	6	7
MAT:32.2.2.1	Não	A-	ni-	ca,	te es-	con-	das,
MAT:32.2.2.2	a-	pa-	re-	ce	sem-	pre,	
MAT:32.2.2.3	que o	ser	bem	pa-	re-	ci-	da
MAT:32.2.2.4	di-	ssó	de-	pen-	de.		

O motivo para a falha da heurística está no segundo verso, que de fato possui 5 sílabas e não 4. É possível que Chociay tenha tido acesso a alguma variante do verso ou que tenha levado em consideração algum fenômeno ignorado pelo autor da presente. Chociay menciona o poema que contém essa estrofe de forma explícita no texto à p. 58.

Os versos octossílabos encontrados pelo sistema, mas nulos na contagem de Chociay, se devem a duas fontes distintas: *a)* os versos cujo texto provavelmente contém erros na edição de James Amado e que por isso foram marcados como inescandíveis durante a revisão pontual; *b)* versos em poemas com redondilhos quebrados, que o sistema errou ao não classificá-los como heptassílabos.

Os decassílabos não italianos se devem, de acordo com Chociay, a três poemas que contém bordões que se destacam ritmicamente. Os dois bordões com esquema 5-10 são, na primeira vez em que aparecem:

	1	2	3	4	5	6	7	8	9	10	11
MAT:1.3.1.11	es-	ta é	a	jus-	ti-	ça,	que	man-	da El-	Rei.	
MAT:8.5.1.11	Não	vo-	s es-	pan-	teis,	que in-	da	lá	vem	mais.	

O terceiro bordão, em 4-7-10, é o seguinte:

	1	2	3	4	5	6	7	8	9	10	11
MAT:5.25.1.9	às	ma-	ra-	vi,	ma-	ra-	vi,	ma-	ra-	vi-	ilhas.

O sistema encontrou 14 casos a mais para o esquema 5-10. Entre os versos excedentes em esquema 5-10 encontrados pelo sistema estão:

	1	2	3	4	5	6	7	8	9	10	11
MAT:5.3.1.4	Da-	r um	di-	a	lei-	s à	su-	a	co-	rren-	te.
MAT:7.2.4.3	Por	lo-	gra-	r e-	ter-	na	vi-	da a	tua	mor-	te.
MAT:18.4.2.5	Se-	res	mes-	tre	vós	na	ve-	lha-	ca-	ri-	a
MAT:39.2.3.2	Es-	ta a-	ca-	de-	mia,	em	que	tan-	to a-	cer-	ta,

É possível que tais versos contenham erros e que Chociay tenha tido acesso a versões em que o ritmo está mais consistente com a prática gregoriana. O segundo verso pode também ser lido com uma sístole em *academia*, o que o tornaria sáfico. Nem todos os versos com esquema 5-10 são justificáveis como os acima; alguns foram escandidos incorretamente, como o seguinte:

	1	2	3	4	5	6	7	8	9	10	11
MAT:17.5.1.1	Che-	gan-	do à	Ca-	jaf-	ba,	vi	An-	to-	ni-	ca,

Seria preferível fazer uma sinalefa entre *vi* e *Antonica* do que a sinérese em *Cajaíba*, o que produziria um esquema 2-6-10 em vez do 2-5-7-10 calculado pelo sistema.

Os 4 casos a mais de decassílabos com padrão 4-7-10 são estes:

	1	2	3	4	5	6	7	8	9	10	11
MAT:2.18.3.3	Da	des-	ven-	tu-	ra	de uns,	d'ou-	tros	ven-	tu-	ra.
MAT:3.7.8.5	Pa-	ra	li-	vrar-	me	de um,	e ou-	tro	tro-	pe-	ço,
MAT:5.27.1.4	Mas	sa-	be as	ciên-	cias,	que	nun-	ca a-	pren-	deu.	
MAT:30.14.4.1	Se i-	ma-	gi-	nas-	tes,	que o	não	sen-	ti-	ri-	a,

Os dois primeiros versos não foram escandidos corretamente pelo sistema, que deixou *uns* e *um* átonos, o que fez com que as heurísticas encontrassem o esquema rítmico incorreto. Os outros dois podem ser pronunciados de forma heroica se o pronome relativo *que* receber um acento, o que seria bastante aceitável dada a grande quantidade de versos heptassilábicos em Gregório que exigem que *que* final seja tônico.

O dodecassílabo encontrado pelo sistema é a escansão correta de um verso incorreto:

	1	2	3	4	5	6	7	8	9	10	11	12	13
MAT:22.17.2.2	Ma-	s ei-	lo	vai,	tu-	do	o	que é	de a-	mo-	ir, o-	bri-	ga,

Este verso pertence a um soneto decassilábico e foi marcado como inescandível na revisão pontual.

Por fim, o erro a que se refere a Tabela 6.16 deve-se a um verso com uma palavra de grafia questionável, que o sistema não soube tratar:

MAT:4.25.1.2 da Ilha da Madeira Alz.

**Discussão.** A contagem de Chociay é mais criteriosa que a junção do Aoi-dos com algumas heurísticas. Ao encontrar um verso que aparentemente tem ritmo 5-10, as heurísticas classificam-no como tal, não possuindo a sofisticação de indagar se o verso não está talvez incorreto e buscar uma versão alternativa em outra edição. O mesmo acontece com os octossílabos: mesmo que o *corpus* marque-os como inescandíveis devido a prováveis erros na edição, o sistema os conta como octossílabos e não heptassílabos, mesmo quando os poemas em que eles se encontram são isométricos. O nível de análise necessário para que os fenômenos de sinafia, compensação, etc. sejam levados em conta durante a análise automática está além das possibilidades do sistema atualmente, o que faz com que versos sejam classificados como trissílabos ou tetrassílabos em vez de redondilhos quebrados.

Por outro lado, o sistema opera de forma robótica e não comete erros de omissão. Se considerarmos que Chociay provavelmente deixou de contar 1.272 versos, então os erros de classificação do sistema e das heurísticas passam a ser proporcionalmente bem menores. O sistema pode ter classificado 18 versos a mais como decassílabos de esquema rítmico 5-10 e 4-7-10, mas contou todos os versos de todos os sonetos, inclusive os 21 que Chociay aparentemente pulou.

Outra vantagem de uma abordagem automática é que os erros e os acertos cometidos por ele são bastante sistemáticos e objetivos. Um humano sempre analisará melhor um verso, mas ao analisar mais de 30 mil, o humano cansa e comete erros que são difíceis de quantificar. O sistema, por outro lado, comete erros que podem via de regra ser localizados e examinados. Num trabalho de contagem manual, qualquer critério que não tenha sido estabelecido desde o princípio requer que a análise comece novamente, ao passo que o sistema pode ser modificado e a análise inteira pode ser feita em poucos minutos. Se Chociay fosse contar o número de sinalefas feitas por Gregório de Matos, por exemplo, precisaria refazer todo o trabalho, o que levaria bastante tempo; o sistema, se não fizesse a coleta dessa informação automaticamente, poderia ser modificado para fazê-lo e o resultado estaria, novamente, pronto em poucos minutos.

## 6.5. Caso de uso: os decassílabos do *Uruguai*

Rogério Chociay publicou um artigo (73) em que classifica os versos do *Uruguai*, de Basílio da Gama, de acordo com o esquema rítmico que eles utilizam, além de comparar esta obra a outras de várias épocas, também com relação aos esquemas rítmicos. Este caso de uso busca definir algumas heurísticas para computar os esquemas rítmicos de decassílabos a partir da escansão do Aoidos e comparar os resultados produzidos de forma automática com os manualmente obtidos por Chociay.

**O esquema rítmico.** O esquema rítmico de um verso é obtido a partir das sílabas que nele estão acentuadas. A tarefa é complexa porque o acento a ser considerado é o icto, o acento do verso, não o das palavras individuais. Palavras átonas podem tornar-se tônicas, por exemplo, ou várias sílabas tônicas podem aparecer em sequência, caso em que algumas geralmente perdem seu acento.

Chociay, em seu artigo, separa sempre os heroicos dos sáficos; os primeiros com 6ª sílaba dominante e os últimos com 4ª. Há inúmeros ca-

sos, contudo, em que tanto a 4ª como a 6ª sílaba são acentuadas. Tais casos não são separados por ele em uma terceira classe; Chociay marca-os ou como heroicos ou como sáficos segundo critérios sintáticos e semânticos. Deste modo, um esquema rítmico não define completamente o tipo do verso, heroico ou sáfico.

**Heurísticas.** Algumas heurísticas foram adicionados ao Aoidos para tentar reproduzir os esquemas rítmicos e a classificação em tipos de Chociay. É importante observar que nem todas as análises presentes na literatura são feitas com os mesmos critérios. Britto (69), por exemplo, ao estabelecer os esquemas rítmicos de um poema decassilábico, afirma que o esquema 2-4-6-8-10 não é nem heroico, nem sáfico — mas jâmbico puro. O artigo em que Chociay analisa o Uruguai, por outro lado, classifica os versos com este esquema rítmico como heroicos ou sáficos, de acordo com certos critérios. Além disso, o próprio esquema rítmico pode estar em discussão. Para Mattoso (126), por exemplo, o seguinte verso de Camões possui esquema rítmico 5-10:

1      2      3      4      5      6      7      8      9      10      11  
CAM.1.1.2 Que|da o-|ci-|den-|tal|prai-|a|Lu-|si-|ta-|na,

Chociay, no artigo do Uruguai, não fornece o esquema rítmico desse verso de forma explícita, mas ele é contado de forma agregada nas tabelas — em que não aparece nenhum esquema rítmico com a 5ª sílaba acentuada. As heurísticas aqui descritas são, portanto, em alguma medida específicas aos critérios que Chociay em seu artigo sobre o Uruguai.

O Aoidos normalmente trabalha com a distinção entre vogais átonas e tônicas, mas para este experimento a distinção foi ampliada para cinco níveis: sílabas átonas, fracas, médias, fortes e fortíssimas. Esta distinção se aplica apenas durante a produção do esquema rítmico e da classificação do verso, não afetando a escansão. A categorização da tonicidade das sílabas se dá com base na classificação inicial binária do Aoidos e com base no texto do verso, numa tentativa de capturar de forma rudimentar a sintaxe do verso.

Todas as sílabas átonas detectadas pelo Aoidos permanecem como átonas. As sílabas inicialmente classificadas como tônicas tornam-se fracas se as palavras a que pertencem estiverem numa lista de 76 formas que incluem preposições como *para* e *entre*, contrações como *pela* e *numa*, pronomes possessivos como *meu* e *seu*, demonstrativos dissilábicos como *este* e *essa*. Tornam-se médias as sílabas inicialmente detectados como tônicas se fizerem parte de uma palavra que figure entre 83 vocábulos, que incluem

demonstrativos dissilábicos como *aquele* e *aquela* — bem como suas contrações com as preposições *a*, *em* e *d* —, pronomes pessoais como *eu* e *tu*, além de algumas conjunções e advérbios. As sílabas tônicas que não se encaixarem em nenhum desses dois casos são categorizadas como fortes. A única exceção é a 10ª sílaba, que é considerada sempre forte.

A pontuação do verso original é examinada para intensificar algumas sílabas não-átonas. Observando-se a silabação do texto do verso de acordo com a escansão, procura-se por um caractere de pontuação, como o ponto, a vírgula, etc., no fim de cada sílaba não-átona; alternativamente, se houver continuação da palavra para as próximas sílabas, são consideradas ainda duas sílabas a mais na busca pela pontuação. O objetivo é encontrar sílabas como a 4ª dos dois versos seguintes:

	1	2	3	4	5	6	7	8	9	10	11
GAM.4.1.234	Eu	mes-	mo os	vi,	que	de-	scem	do al-	to	mon-	te,
GAM.3.1.204	Vi-	si-	o-	ná-	ria,	su-	pers-	ti-	ci-	o-	sa,

Estas sílabas encontram-se sintaticamente destacadas e por isso tem sua tonicidade ampliada pelo sistema. A promoção é feita do seguinte modo: fracas tornam-se médias; médias, fortes; fortes, fortíssimas.

Para os graus de tonicidade fortíssimo, forte e médio, nesta ordem, são examinadas as sílabas nas posições 6, 4 e 8, nesta ordem. Sempre que a sílaba na dada posição possuir o dado grau de tonicidade, as duas sílabas ao redor da considerada são tornadas átonas. Para os mesmos graus, são então examinadas todas as sílabas, do fim para o início: quando a sílaba tiver aquele grau de tonicidade, então a sílaba precedente é marcada como átona. As sílabas fracas remanescentes são, por fim, reduzidas a átonas.

A posição de todas as sílabas não-átônicas são, neste estágio, incluídas no esquema rítmico de forma indistinta, mas o grau de tonicidade é utilizado ainda na classificação dos versos nos tipos heroico, sáfico e outro. Se as posições 4 e 6 foram não-átônicas, é observado o seu grau de tonicidade: se a 4ª sílaba for mais forte que a 6ª, o verso é sáfico; caso contrário é heroico. Se a 6ª for não-átônica sem que a 4ª o seja, o verso é heroico. Se a 4ª for não-átônica sem que a 6ª e a 7ª o sejam, o verso é sáfico. Se não for categorizado por nenhuma dessas condições, o verso é classificado como outro.

**Versos concretos.** Antes de mencionar números agregados, Chociay exemplifica 16 versos heroicos, 12 sáficos e 3 excepcionais, juntamente com o esquema rítmico de cada um. A Tabela 6.17 compara o resultado produzido pelo Aoidos com a análise manual de Chociay. As duas análises, em

geral, concordam, excetuados três casos. O verso GAM.1.1.111 foi escandido de forma possivelmente incorreta pelo sistema:

1    2    3    4    5    6    7    8    9    10   11  
GAM.1.1.111 O|ros-|to es-|mo-|re-|ci-|do|e os|frios|bra-|ços,

Da maneira como está escandido, o esquema rítmico produzido pelo sistema está correto; questionável é a escansão. A segunda discordância é um erro de Chociay, que considerou a segunda sílaba tônica no seguinte verso:

1    2    3    4    5    6    7    8    9    10   11  
GAM.2.1.216 Co-|brem|as|tro-|pas|de-|ca-|va-|le-|ri-|a,

O verso está escrito dessa maneira e com a referência correta no artigo de Chociay. Por fim, o sistema classificou como heroico um verso que Chociay afirma ser sáfico:

1    2    3    4    5    6    7    8    9    10   11  
GAM.1.1.143 Al-|ta-|s em-|pre-|sas|dig-|nas|de-|me-|mó-|ria.

O esquema rítmico 1-4-6-10, encontrado pelas duas partes, é resolvido por Chociay dando mais importância à palavra *empresas* que a *dignas*.

**Contagem no *Uruguai*.** Chociay encontrou manualmente o esquema rítmico de todos os 1.377 versos do *Uruguai* e os classificou em heroicos, sáficos e excepcionais. Os resultados dessa análise manual e também da automática feita pelo Aoidos constam das Tabelas 6.18 e 6.19, que agrupam, respectivamente, os heroicos e os sáficos. O Aoidos encontrou ainda 10 versos excepcionais, enquanto Chociay encontrou 13 mas não os agrupou segundo seu esquema rítmico.

É difícil apontar a origem das diferenças encontradas entre os números automáticos e manuais, pois estes últimos estão disponíveis apenas de forma sintética. Não existe, de fato, garantia nenhuma de que quando a diferença é zero as análises computaram naquele esquema rítmico os mesmos versos. O valor absoluto baixo das diferenças é evidência de que as análises são similares, mas discutir casos específicos de esquemas rítmicos tem pouca utilidade, dado que não se sabe quais versos a análise manual inclui em cada esquema.

O que pode ser comparado são tendências mais gerais. A Tabela 6.20 mostra o percentual de cada tipo de verso ao longo do *Uruguai* segundo as classificações manual e automática. A maior diferença, em pontos percentuais, está no Canto V, em que o Aoidos contou 6 pontos a mais de heroicos do que Chociay e, inversamente, 6,7 pontos a menos de sáficos. Este é o canto com a menor quantidade de versos: ele possui 150 versos, enquanto os quatro primeiros contêm 244, 365, 329 e 289, res-



Tabela 6.17.: Versos do *Uruguai* explicitamente classificados por Chociay, c, em comparação com a classificação do Aoidos, A. Os versos estão classificados segundo o tipo, H = heroico, S = sáfico e O = outro, e segundo o esquema rítmico. Os itens sublinhados indicam discrepância.

Código	Tipo		Esquema		Texto
	A	C	A	C	
GAM.5.1.45	H	H	6-10	6-10	E da navegação e do comércio.
GAM.1.1.30	H	H	1-6-10	1-6-10	Já por dilatadíssimos caminhos
GAM.1.1.197	H	H	2-6-10	2-6-10	De bárbaros o número infinito
GAM.1.1.32	H	H	3-6-10	3-6-10	Conduzir os petrechos para a guerra.
GAM.2.1.77	H	H	4-6-10	4-6-10	Às sitiadoras tropas castelhanas
GAM.1.1.52	H	H	6-8-10	6-8-10	E lhe determinou lugar e tempo
GAM.5.1.135	H	H	1-3-6-10	1-3-6-10	Cai a infame República por terra.
GAM.1.1.2	H	H	1-4-6-10	1-4-6-10	Lagos de sangue tépidos e impuros
GAM.2.1.82	H	H	1-6-8-10	1-6-8-10	Toda a navegação do largo rio,
GAM.1.1.5	H	H	2-4-6-10	2-4-6-10	O rouco som da irada artilheria.
GAM.1.1.111	H	H	<u>2-6-10</u>	<u>2-6-8-10</u>	O rosto esmorecido e os frios braços,
GAM.1.1.8	H	H	<u>3-6-8-10</u>	<u>3-6-8-10</u>	Dos decretos reais lavou a afronta.
GAM.3.1.179	H	H	4-6-8-10	4-6-8-10	Da enternecida esposa abranda o peito
GAM.1.1.59	H	H	1-3-6-8-10	1-3-6-8-10	Tudo nota de parte e tudo observa
GAM.1.1.110	H	H	1-4-6-8-10	1-4-6-8-10	Ninfas do amor, que vistes, se é que vistes,
GAM.3.1.223	H	H	2-4-6-8-10	2-4-6-8-10	O rio, a praia o vale e os montes onde
GAM.3.1.204	S	S	4-10	4-10	Visionária, supersticiosa,
GAM.4.1.152	S	S	1-4-10	1-4-10	Lá reclinada, como que dormia,
GAM.2.1.216	S	S	<u>1-4-10</u>	<u>2-4-10</u>	Cobrem as tropas de cavalaria,
GAM.1.1.162	S	S	4-6-10	4-6-10	Como sabeis, neste ângulo da terra,
GAM.1.1.141	S	S	4-8-10	4-8-10	Arrebatado de furor divino
GAM.1.1.143	<u>H</u>	<u>S</u>	1-4-6-10	1-4-6-10	Altas empresas dignas de memória.
GAM.1.1.1	S	S	1-4-8-10	1-4-8-10	Fumam ainda nas desertas praias
GAM.5.1.131	S	S	2-4-6-10	2-4-6-10	Chorosas mães, e filhos inocentes,
GAM.1.1.68	S	S	2-4-8-10	2-4-8-10	E a praça e as ruas da cidade errante.
GAM.2.1.13	S	S	4-6-8-10	4-6-8-10	Galopar, a quem primeiro os segue
GAM.1.1.54	S	S	1-4-6-8-10	1-4-6-8-10	Juntos enfim, e um corpo do O à vista,
GAM.1.1.116	S	S	2-4-6-8-10	2-4-6-8-10	Também te viu naquele dia o campo,
GAM.2.1.263	O	O	1-4-7-10	1-4-7-10	Treme, e o cavalo aos seus volta, e pendente
GAM.3.1.162	O	O	3-8-10	3-8-10	Das trombetas lho arrebatou dos braços
GAM.4.1.107	O	O	3-10	3-10	Que combate desordenadamente.

pectivamente. Os erros absolutos mostrados nas Tabelas 6.18 e 6.19, de 9 heroicos a mais e 10 sáficos a menos, pesam bastante sobre a pequena quantidade de versos do canto. Um gráfico das contagens, como aquele da Figura 6.1, mostra que as tendências na obra, apontadas pela análise manual de Chociay, são plenamente visíveis também na análise automática.

**Contagem em sete *corpora*.** Chociay também compara os 100 primeiros versos do *Uraguai* aos 100 primeiros de seis outras obras de épocas variadas. As contagens da análise automática e a diferença com relação à manual estão à mostra nas Tabelas 6.21 e 6.22, para os versos heroicos e sáficos, respectivamente. Também aqui há variações dentro de cada tipo de verso, alguns esquemas rítmicos recebendo mais versos do que outros, mas em geral os números da análise automática estão próximos aos da manual. Os percentuais de sáficos e heroicos nos *corpora*, que constam da Tabela 6.23, são testemunho de que a proporção de cada tipo está bem preservada na análise manual: a maior diferença é de 3 pontos percentuais (ou 3 versos), no caso *corpus* CAM100. Um gráfico das contagens de heroicos e sáficos, como o da Figura 6.2, corrobora a hipótese de que a análise automática, se não exata nos detalhes, globalmente fornece informações estilísticas compatíveis com a manual.

Tabela 6.18.: Contagem de versos heroicos do *Uruguai*, classificados em seus esquemas rítmicos e divididos por canto. As colunas são A = resultado do Aoidos, c = números de Chociay, D = diferença.

Esquema	Canto I			Canto II			Canto III			Canto IV			Canto V		
	A	C	D	A	C	D	A	C	D	A	C	D	A	C	D
1-3-6-10	13	13	0	19	10	9	9	8	1	14	11	3	7	8	-1
1-3-6-8-10	10	11	-1	10	10	0	9	6	3	10	10	0	2	3	-1
1-4-6-10	9	8	1	11	12	-1	7	9	-2	12	11	1	6	4	2
1-4-6-8-10	6	6	0	11	14	-3	4	3	1	8	8	0	5	4	1
1-6-10	9	8	1	6	6	0	4	6	-2	5	6	-1	2	2	0
1-6-8-10	5	4	1	6	4	2	3	2	1	1	1	0	4	4	0
2-4-6-10	16	13	3	24	27	-3	11	11	0	13	12	1	18	14	4
2-4-6-8-10	10	8	2	22	22	0	19	11	8	12	19	-7	6	7	-1
2-6-10	33	30	3	36	42	-6	24	27	-3	35	32	3	13	14	-1
2-6-8-10	14	19	-5	24	26	-2	22	22	0	29	30	-1	11	10	1
3-6-10	27	26	1	49	50	-1	34	36	-2	24	25	-1	17	19	-2
3-6-8-10	18	22	-4	35	42	-7	29	36	-7	22	25	-3	13	13	0
4-6-10	2	3	-1	11	8	3	16	11	5	10	7	3	6	3	3
4-6-8-10	3	3	0	9	4	5	10	7	3	11	7	4	4	1	3
6-10	3	1	2	4	0	4	1	1	0	0	0	0	1	1	0
6-8-10	5	2	3	1	1	0	0	0	0	1	0	1	1	0	1
	183	177	6	278	278	0	202	196	6	207	204	3	116	107	9

Tabela 6.19.: Contagem de versos sáficos do *Uruguai*, classificados em seus esquemas rítmicos e divididos por canto. As colunas são A = resultado do Aoidos, C = números de Chociay, D = diferença.

Esquema	Canto I			Canto II			Canto III			Canto IV			Canto V		
	A	C	D	A	C	D	A	C	D	A	C	D	A	C	D
1-4-10	1	1	0	1	0	1	0	0	0	1	1	0	0	0	0
1-4-6-10	2	4	-2	6	4	2	0	0	0	1	3	-2	0	0	0
1-4-6-8-10	2	1	1	3	3	0	2	1	1	2	1	1	0	2	-2
1-4-8-10	15	15	0	13	12	1	25	26	-1	13	15	-2	5	4	1
2-4-10	0	0	0	3	3	0	0	0	0	1	1	0	0	1	-1
2-4-6-10	2	1	1	1	0	1	5	1	4	2	2	0	2	6	-4
2-4-6-8-10	4	6	-2	4	2	2	7	11	-4	3	5	-2	0	2	-2
2-4-8-10	19	22	-3	32	34	-2	57	63	-6	39	41	-2	16	17	-1
4-10	0	0	0	2	2	0	4	4	0	0	0	0	0	0	0
4-6-10	1	2	-1	2	1	1	2	0	2	2	0	2	1	2	-1
4-6-8-10	2	0	2	2	4	-2	0	2	-2	1	2	-1	0	0	0
4-8-10	13	13	0	16	17	-1	19	22	-3	16	11	5	9	9	0
	61	65	-4	85	82	3	121	130	-9	81	82	-1	33	43	-10

Tabela 6.20.: Percentual de versos H = heroicos, s = sáfcos e O = outros do *Uraguai* de acordo com cada canto. As colunas são A = resultado do Aoi-dos, C = números de Chociay, D = diferença, em pontos percentuais.

	Canto I			Canto II			Canto III			Canto IV			Canto V		
T	A	C	D	A	C	D	A	C	D	A	C	D	A	C	D
H	75,0	72,5	<b>2,5</b>	76,2	76,2	<b>0,0</b>	61,4	59,6	<b>1,8</b>	71,6	70,6	<b>1,0</b>	77,3	71,3	<b>6,0</b>
s	25,0	26,6	<b>-1,6</b>	23,3	22,5	<b>0,8</b>	36,8	39,5	<b>-2,7</b>	28,0	28,4	<b>-0,3</b>	22,0	28,7	<b>-6,7</b>
O	0,0	0,8	<b>-0,8</b>	0,5	1,4	<b>-0,8</b>	1,8	0,9	<b>0,9</b>	0,3	1,0	<b>-0,7</b>	0,7	0,0	<b>0,7</b>

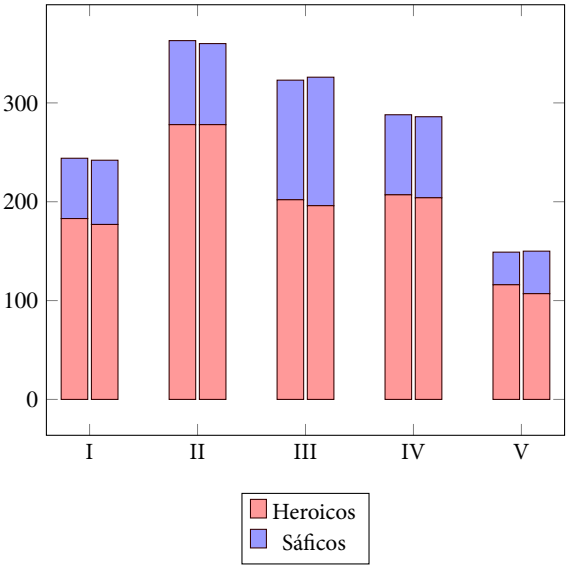


Figura 6.1.: Contagem de heroicos e sáfcos dos cantos do *Uraguai*. Em cada canto à esquerda está a contagem automática e à direita a de Chociay.

Tabela 6.21.: Contagem dos versos heroicos de sete *corpora*, classificados de acordo com o esquema rítmico. As colunas são A = resultado do Aoidos, c = números de Chociay, D = diferença.

Esquema	CAM100		GAM100		DUR100		MAN100		DIA100		OLI100		LIM100	
	A	D	A	D	A	D	A	D	A	D	A	D	A	D
1-3-6-10	2	-2	7	-2	4	-2	3	0	2	-1	4	0	3	2
1-3-6-8-10	6	2	3	-1	10	1	3	-2	1	0	2	2	4	2
1-3-8-10	0	0	0	0	0	0	0	0	0	0	0	-1	0	0
1-4-6-10	9	0	4	2	6	0	4	0	5	2	3	-2	2	-2
1-4-6-8-10	6	1	2	0	4	1	1	0	1	-1	2	0	2	0
1-6-10	2	0	5	1	0	-1	1	0	0	0	4	1	1	0
1-6-8-10	0	-1	1	0	2	1	1	-1	2	0	1	0	2	0
2-4-6-10	13	1	5	0	4	1	6	2	10	-2	5	1	5	-1
2-4-6-8-10	4	-2	4	0	5	1	6	-3	5	0	3	0	5	0
2-6-10	14	-3	12	2	9	-4	12	-1	9	-1	5	-2	17	-2
2-6-8-10	7	-6	3	1	6	1	13	4	15	-1	6	1	8	-2
2-8-10	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0
3-6-10	11	1	14	0	11	2	9	2	6	0	12	0	12	3
3-6-8-10	5	0	13	0	11	-2	12	0	5	1	8	0	9	-2
4-6-10	5	2	1	-1	3	0	0	-1	3	0	6	0	4	2
4-6-8-10	5	2	1	0	1	0	3	2	2	1	1	-1	2	0
6-10	2	0	0	0	0	0	0	0	0	0	2	1	2	0
6-8-10	2	2	1	0	0	0	0	0	0	0	1	0	2	1
	93	-3	76	1	76	-1	74	2	66	-2	65	-1	80	1

Tabela 6.22.: Contagem dos versos sáficos de sete *corpora*, classificados de acordo com o esquema rítmico. As colunas são A = resultado do Aoidos, C = números de Chociay, D = diferença.

Esquema	CAM100		GAM100		DUR100		MAN100		DIA100		OLI100		LIM100	
	A	D	A	D	A	D	A	D	A	D	A	D	A	D
1-4-10	0	0	0	0	1	1	0	0	0	0	2	2	0	-1
1-4-6-10	0	0	1	-1	1	0	1	0	0	0	0	0	1	1
1-4-6-8-10	1	0	1	0	1	0	1	1	1	1	1	0	2	1
1-4-8-10	0	0	8	0	5	-1	5	0	4	0	8	-2	1	-1
2-4-10	1	0	0	0	0	0	0	0	0	0	1	-1	1	0
2-4-6-10	3	3	0	0	0	0	4	2	3	1	1	0	2	1
2-4-6-8-10	1	-1	1	0	0	-1	2	-2	1	-1	2	-2	5	-3
2-4-8-10	0	0	9	0	13	0	10	-1	22	0	16	3	2	-1
4-10	0	0	0	0	0	0	0	0	0	0	0	0	1	1
4-6-10	0	0	0	0	0	0	0	-2	1	0	1	1	2	1
4-6-8-10	1	1	1	0	0	0	0	-1	0	0	0	0	2	1
4-8-10	0	0	3	0	2	1	3	1	2	1	3	1	1	0
	7	3	24	-1	23	0	26	-2	34	2	35	2	20	0

Tabela 6.23.: Percentual de versos H = heroicos, s = sáficos e o = outros entre os *corpora*. As colunas são A = resultado do Aoidos, C = números de Chociay, D = diferença, em pontos percentuais.

E	CAM100			GAM100			DUR100			MAN100			DIA100			OLI100			LIM100		
	A	C	D	A	C	D	A	C	D	A	C	D	A	C	D	A	C	D	A	C	D
H	93	96	-3	76	75	1	76	77	-1	74	72	2	66	68	-2	65	66	-1	80	79	1
s	7	4	3	24	25	-1	23	23	0	26	28	-2	34	32	2	35	34	1	20	20	0
o	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	-1

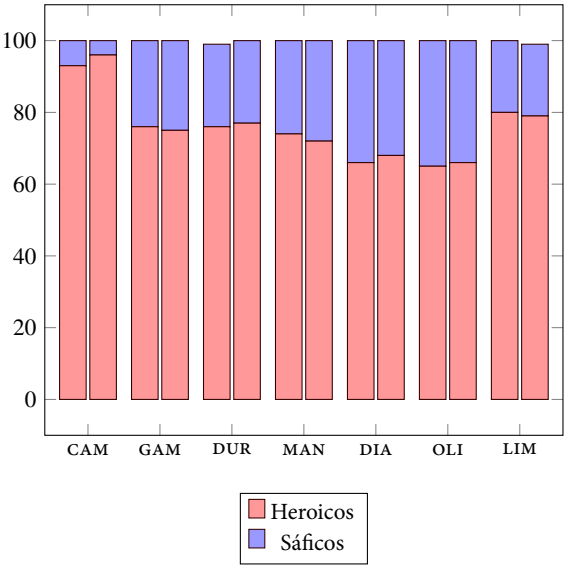


Figura 6.2.: Contagem de heroicos e sáficos dos *corpora*. Em cada *corpus* à esquerda está a contagem automática e à direita a de Chociay.



## 7. CONCLUSÃO

1	2	3	4	5	6	7	8	9	10	11
São	ri-	mas	de	ta-	re-	lo, a-	tro-	pe-	la-	das,
Sem	me-	tro,	sem	ca-	dên-	cia e	sem	bi-	to-	la,
Que	for-	mam	no	pa-	pel	um	zi-	gue-	za-	gue,
Co-	mo os	pa-	ssos	de	ren-	go	man-	qui-	to-	la.

Luís Gama

Esta tese descreve o Aoidos, um sistema completo de escansão automática de versos em português. O sistema começa do próprio texto, localiza o acento e faz a divisão silábica dos vocábulos, gera alternativas de escansão para cada verso, determina a quantidade de sílabas de cada verso e escolhe a melhor alternativa através de critérios rítmicos. Os experimentos avaliaram manualmente a escansão produzida pelo sistema para 1.080 dos 107.438 versos processados, resultando na acurácia de 99,0%.

Este é o primeiro sistema de escansão automática para a língua portuguesa que busca ajustar o verso à métrica do poema. O único outro sistema disponível na literatura (124) opera de maneira completamente determinística e possui uma taxa de acerto de 82,2%, avaliada com base em 197 versos. O Aoidos, desta forma, coloca o português entre as línguas que possuem um sistema capaz de escandir versos em grande escala e com boa acurácia.

**Objetivos alcançados.** O objetivo geral da tese, exposto na introdução, foi plenamente atingido. De um lado, a proposta do sistema está formalizada por sua descrição pormenorizada feita no Capítulo 5. De outro, o Capítulo 6 avaliou o sistema através de dois experimentos e dois casos de uso. Um experimento, o principal, manualmente examinou com o auxílio de especialistas uma pequena proporção (1% ou 1.080 versos) de um grande *corpus*, obtendo a taxa de 99,0% de acerto; o outro experimento avaliou uma grande proporção (23,6% ou 1.554 versos) da obra de um só poeta, Augusto dos Anjos, com taxa de acerto de 99,3%. Os casos de uso buscaram reproduzir resultados quantitativos disponíveis na literatura para a poesia de Gregório de Matos, Basílio da Gama e outros; os resultados

mostram que o Aoidos consegue gerar análises muito similares às aquelas feitas manualmente. Com estes experimentos e casos de uso, a hipótese de que é possível escandir versos em português de forma automática fica demonstrada.

Os objetivos específicos foram também cumpridos. Seguindo-os item a item como foram apresentados na introdução:

- Os *corpora* poéticos utilizados nos experimentos e casos de uso, que juntos somam 113.982 versos em português, foram descritos em detalhes no Capítulo 4. Os *corpora* foram adaptados majoritariamente a partir de edições disponíveis na Literatura Digital<sup>1</sup> e sofreram revisões e conferências ao serem analisados pelo Aoidos. Os arquivos estão disponibilizados *online* no seguinte endereço:  
<https://github.com/adiel-mittmann/poemas>
- O algoritmo de localização da vogal tônica é original e foi descrito em minúcias na Seção 5.3 do Capítulo 5 e validado com um experimento em que 40.000 palavras foram manualmente verificadas, com uma taxa de acerto média de 99,3%. Este é um algoritmo que, apesar de ser fundamentalmente importante para a escansão automática, não é de maneira nenhuma específico a ela.
- O algoritmo de divisão silábica, também original, encontra-se proposto na Seção 5.4 do Capítulo 5. Também trata-se de um algoritmo que pode ser empregado em aplicações não relacionadas à poesia.
- As 159 regras prosódicas que o Aoidos utiliza para gerar alternativas de pronúncia estão descritas no Apêndice B e sua maneira de utilização na Seção 5.5.
- O algoritmo de detecção da métrica de poemas foi criado especialmente para o Aoidos e está plenamente descrito na Seção 5.8 do Capítulo 5. O algoritmo classifica o poema em quatro classes, conforme o número de metros presente, e pode trabalhar tanto no nível do poema quanto da estrofe.
- O algoritmo que busca a melhor alternativa rítmica para um verso foi proposto de forma original nesta tese e está apresentado na Seção 5.9. Ele pode trabalhar de forma independente ou utilizar um catálogo de ritmos, que informa ao sistema as expectativas que um

---

<sup>1</sup><http://www.literaturabrasileira.ufsc.br/>

leitor de poesia humano normalmente tem a respeito dos versos em português.

- Por fim, os experimentos e casos de uso do Capítulo 6 avaliam os resultados finais obtidos pelo conjunto de algoritmos do Aoidos e mostram que eles são aceitáveis, o que indiretamente auxilia na validação de cada um dos algoritmos.

**Contribuições.** A contribuição central desta tese é a apresentação e avaliação do primeiro sistema de escansão automática de versos em português com boa acurácia. Como contribuições secundárias destacam-se os algoritmos de localização da vogal tônica de vocábulos, de divisão silábica, de análise métrica e de análise rítmica, todos originais. Também é relevante a disponibilização dos *corpora* utilizados nos experimentos, que contêm mais de 110.000 versos.

**Limitações.** O sistema atualmente depende de um *corpus* bem preparado para escandir os versos corretamente. O número de erros encontrados nas edições digitais e impressas relatados no Capítulo 4 é evidência de que, sob um ângulo prático, é preciso considerar que os versos *sempre* conterão erros. Escandir um verso cujo texto está incorreto é, evidentemente, impossível; a limitação do sistema está na influência negativa que um único verso pode ter sobre a escansão de todos os versos do poema em que ele se encontra. Se a análise métrica do poema não consegue encaixar o verso incorreto em suas hipóteses, então corre-se o risco de que a métrica do poema seja interpretada de forma errônea, o que causa um efeito cascata: se o sistema decide que o poema tem dois metros em vez de um, agora vários versos podem se encaixar no metro adicional, errado. A análise métrica contém parâmetros de limiar que podem ser ajustados, mas o impacto de um verso incorreto num poema pequeno, como um soneto, continua grande.

Mesmo que o *corpus* não contenha nenhum erro, ainda assim ele não está pronto para ser submetido ao sistema. Os estrangeirismos presentes no texto, bem como a expansão em palavras por extenso de números e abreviaturas, precisam ser marcados com sua pronúncia na ortografia usual. A discussão conduzida na Seção 5.1 do Capítulo 5 mostra que esse é um problema difícil de ser resolvido. Em particular, é necessário que os diferentes estágios do sistema cooperem: para saber se *1\$200* deve ser pronunciado com ou sem o nome da moeda, é preciso saber quantas

sílabas o verso deve ter e verificar qual das duas alternativas torna o verso escandível. Mesmo os casos que teoricamente são possíveis de serem resolvidos numa mesma etapa não são fáceis: *VI* pode ser o numeral 6 ou a 1ª pessoa do singular do perfeito do verbo *ver* em maiúsculas; *S.* pode ser *Santo* ou *São*.

Embora os algoritmos de localização da vogal tônica e de divisão silábica apresentem bom desempenho, a transcrição fonética está muito aquém do que pode ser atingido por um sistema baseado em regras. Que ela não tenha sido melhorada é sintoma de que ela não precisou sê-lo. As duas principais fontes de erro, a distinção entre vogais abertas e fechadas em sílaba tônica e a pronúncia da letra *X*, apenas em um caso produziram um problema incontornável: quando a palavra *reflexos* precisou ser pronunciada em quatro sílabas por meio de uma epêntese, o sistema errou ao transcrever o *X* como uma sibilante. De qualquer modo, para futuras análises estilísticas — de aliteração, por exemplo — o sistema precisa de uma transcrição mais fiel à pronúncia.

A análise métrica atualmente não trata versos compostos. Ao encontrar um poema com versos com dois hemistíquios de 5 sílabas, o sistema ou considera erroneamente que eles possuem mais de um metro (por exemplo, 10 e 12 sílabas) ou, o que talvez seja pior, encontra escansões forçadas para todos os versos. Os versos compostos que não podem ser metricamente distinguidos dos simples são, contudo, escandidos corretamente. Resolver este problema requer, no mínimo, que hipóteses adicionais sejam testadas pela análise métrica e que a análise rítmica consiga desprezar sílabas supérfluas no interior do verso.

O sistema como um todo atualmente distingue apenas sílabas átonas e tônicas. Uma nuance maior, como aquela provida pelas heurísticas descritas na Seção 6.5 do Capítulo 6, poderia dar maior sofisticação ao sistema, sobretudo à análise rítmica. Para o Aoidos, é indiferente se uma sílaba que deveria ser tônica é ocupada por uma palavra como *meu* ou *rosto*, ainda que, existindo a opção, é melhor que esta última ocupe uma posição tônica. Por não dispor de uma visão menos binária do acento, o sistema atualmente também não leva em consideração certas noções conhecidas pelos leitores humanos, como a preferência por não deixar duas palavras tônicas próximas uma à outra.

**Trabalho futuro.** Na etapa de localização do acento das palavras, um texto escrito segundo normas ortográficas contemporâneas é essencial. Antes da reforma dos anos 1940, não era possível saber qual vogal acentuar

considerando-se a palavra em isolamento. Palavras com estrutura similar não recebiam o acento na mesma sílaba, como *macio*, que continua sendo escrita da mesma forma, e *ódio*, que hoje escreve-se com agudo: *ódio*. Há até mesmo palavras que se escreviam de forma idêntica mas tinham pronúncia diversa. Por exemplo, *publico* e *practica* podiam tanto ser o substantivo, que hoje escrevemos *público* e *prática*, quanto a 1ª pessoa do singular do presente indicativo do verbo *publicar* e a 3ª do singular do *praticar*. Não tratar a ortografia antiga significa exigir a atualização ortográfica de um *corpus* antes de ser analisado pelo sistema, o que é um obstáculo no tratamento de livros antigos ou em versos modernos que seguem a norma antiga, como a produção poética de Glauco Mattoso Mattoso (125).

A incorporação da análise sintática como uma etapa inicial traria vários benefícios ao sistema. A transcrição fonética poderia incluir regras para, por exemplo, distinguir entre *olho*, a 1ª pessoa do singular do presente indicativo de *olhar*, e *olho*, o substantivo. Uma palavra como *um* poderia receber ou não acento conforme sua utilização como artigo ou como numeral. A análise rítmica, ao passar a considerar níveis de acentuação, poderia tratar de formas diferentes palavras de classes gramaticais distintas. A análise sintática, portanto, poderia beneficiar várias partes do sistema.

As estatística geradas pelo sistema acerca da presença de fenômenos como elisões e síncope baseia-se atualmente na contagem da aplicação das regras prosódicas. Isto significa que apenas as elisões e síncope *não marcadas* no texto são contadas. Para que uma análise estilística mais fiel possa ser feita, é preciso que o sistema conte também os fenômenos marcados no texto. No caso de síncope, fazê-lo requereria localizar apóstrofes dentro de formas como *f'licidade* e contabilizá-los como marcadores do fenômeno. Existem outros casos, todavia, que são mais complexos. Uma sístole marcada no texto utiliza a ortografia regular do português; para saber que *periféria* é uma sístole, é preciso saber que a pronúncia habitual é *periferia*.

O sistema atualmente tem poucas expectativas a respeito da métrica e do ritmo dos poemas que ele analisa. O único conhecimento prévio que pode ser passado ao sistema é o catálogo de ritmos, que diz, por exemplo, que os decassílabos heroicos, de esquema rítmico 6-10, são comuns. O sistema nada assume, porém, a respeito da frequência dos versos. Em poemas pequenos, como quadras avulsas, em que há pouco que se possa tomar como base para encontrar a métrica, o sistema pode facilmente classificar os versos como hendecassílabos por causa de algumas sinalefas com-

plexas; se ele soubesse que os hendecassílabos são bem menos comuns e que quase sempre tem esquema 5-11, talvez não fizesse essa classificação. Outros exemplos de conhecimentos prévios que o sistema poderia opcionalmente utilizar diz respeito à pontuação das regras: movimentos literários diferentes dão importâncias distintas a certos fenômenos. A presença de um hiato, por exemplo, pode ser considerada uma falha grave num período, enquanto é aceitável em outro; uma regra de sinalefa que desfaz o hiato poderia levar isso em conta e ajustar sua pontuação.

A pontuação de penalidade das regras prosódicas utilizadas pelo sistema para a geração de alternativas de pronúncia para os versos é atualmente bastante subjetiva. O fato de que as escansões produzidas pelo sistema sejam majoritariamente corretas indica que, embora subjetivas, as pontuações fazem sentido. Determinar as pontuações de forma mais sistemática permitiria não apenas justificar melhor os resultados produzidos pelo sistema mas também enriquecer as informações estilísticas derivadas da escansão, permitindo uma comparação mais embasada entre os metaplasmos aplicados a diferentes conjuntos de versos.

O tratamento de *corpora* não tratados aumentaria a aplicabilidade do sistema. Além das limitações descritas acima com relação a estrangeirismos e abreviações, um *corpus* não tratado inclui vários elementos que não devem ser incluídos na análise, como títulos, assinaturas, epígrafes, etc. Dependendo da formatação usada na edição digital a ser submetida a uma extração de versos, estes elementos podem ter uma aparência muito similar aos versos. Ademais, a presença de elementos como nomes de personagens ou marcadores de pergunta e resposta, que podem não estar separados visualmente dos versos, também torna o problema mais difícil. Tratar obras dramáticas escritas em verso, por fim, representa um desafio ainda maior: em diálogos, o número de versos emendados, que começam com uma personagem e terminam com outra, é grande.

Devido à presença perene de erros nos poemas, o sistema pode encontrar uma escansão para um verso quando o melhor seria que ele indicasse um erro. Para um verso em que falta uma palavra, por exemplo, o sistema pode produzir uma escansão com vários hiatos que compensam a ausência. A detecção de situações anormais poderia ser feita tanto com base nas regras que foram aplicadas — uma palavra a mais no verso pode exigir sinalefas complexas demais — quanto na pontuação da escansão final com relação à alternativa mais natural encontrada pelo sistema. Assim, ao tratar um *corpus* novo, o usuário teria a possibilidade de ser alertado acerca de escansões suspeitas para que pudesse verificar o verso em busca

de possíveis erros.

A escansão produzida pelo sistema fornece dados estatísticos que caracterizam o estilo dos versos e que são valiosos em si, mas outras maneiras de explorar os resultados podem ser buscadas. Ao longo da pesquisa que culminou nesta tese, foi criado um esquema hierárquico de visualização (130). Com esse esquema, o usuário pode visualizar desde os fonemas dos versos até livros inteiros de poesia. Não se trata de uma proposta isolada; outros autores têm proposto ferramentas com foco na visualização, como é o caso do Poemage (128). Desenvolver visualizações e interfaces gráficas é importante para que o sistema possa ser usado na prática.





## BIBLIOGRAFIA

### Fontes

- 1 ALIGHIERI, D. *Divina Comédia*. Tradução: José Pedro Xavier Pinheiro. Rio de Janeiro: Instituto Profissional Masculino, 1907. Disponível em: <<https://goo.gl/Dhz6TT>>. Acesso em: 15 dez. 2016.
- 2 ALIGHIERI, D. *Divina Comédia*. Tradução: José Pedro Xavier Pinheiro. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/2VC7Zy>>. Acesso em: 15 dez. 2016.
- 3 ANJOS, A. dos. *Obra Completa*. Rio de Janeiro: Nova Aguilar, 1994.
- 4 ANJOS, A. dos. *Obra Completa*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/gnzTTL>>. Acesso em: 15 dez. 2016.
- 5 ARAÚJO FIGUEREDO, J. de. *Poesias*. Florianópolis: Academia Catarinense de Letras, 1966.
- 6 ARAÚJO FIGUEREDO, J. de. *Poesias*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/N4Gk9W>>. Acesso em: 15 dez. 2016.
- 7 CAMÕES, L. de. *Os Lusíadas*. 4ª ed. Lisboa: Ministério dos Negócios Estrangeiros, Instituto Camões, 2000. Disponível em: <<https://goo.gl/qb5bQA>>. Acesso em: 14 dez. 2016.
- 8 CAMÕES, L. de. *Os Lusíadas*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/6S70dJ>>. Acesso em: 14 dez. 2016.
- 9 COSTA, C. M. da. *Obras de Cláudio Manuel da Costa, Arcade Ultramarino, Chamado Glauceste Satúrnio*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/rXEpMv>>. Acesso em: 15 dez. 2016.
- 10 COSTA, C. M. da. *Vila Rica*. Ouro Preto: Tipografia do Universal, 1839. Disponível em: <<https://goo.gl/0NtcNw>>. Acesso em: 15 dez. 2016.

- 11 COSTA, C. M. da. *Vila Rica*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/DJP3rC>>. Acesso em: 15 dez. 2016.
- 12 COSTA, C. M. da; GONZAGA, T. A.; PEIXOTO, A. *A Poesia dos Inconfidentes*. Rio de Janeiro: Nova Aguilar, 1996.
- 13 DIAS, A. G. *Os Timbiras*. Leipzig: F. A. Brockhaus, 1857. Disponível em: <<https://goo.gl/uNbCJ9>>. Acesso em: 15 dez. 2016.
- 14 DIAS, A. G. *Os Timbiras*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/6owqYg>>. Acesso em: 15 dez. 2016.
- 15 GAMA, J. B. da. *O Uruguai*. Lisboa: Régia Oficina Tipográfica, 1769. Disponível em: <<https://goo.gl/tDnbdS>>. Acesso em: 15 dez. 2016.
- 16 GAMA, J. B. da. *O Uruguai*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/6vH0x6>>. Acesso em: 15 dez. 2016.
- 17 GONZAGA, T. A. *Cartas Chilenas*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/W585Ty>>. Acesso em: 15 dez. 2016.
- 18 LIMA, J. de. *Invenção de Orfeu*. Rio de Janeiro: Livros de Portugal, 1952.
- 19 MAGALHÃES, D. J. G. de. *Suspiros Poéticos e Saudades*. Rio de Janeiro: João Pedro da Veiga, 1836. Disponível em: <<https://goo.gl/qSrUDr>>. Acesso em: 15 dez. 2016.
- 20 MAGALHÃES, D. J. G. de. *Suspiros Poéticos e Saudades*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/ZyqdEP>>. Acesso em: 15 dez. 2016.
- 21 MATOS, G. de. *Obra Poética*. Edição: James Amado. 3ª edição. Rio de Janeiro: Record, 1992.
- 22 MATOS, G. de. *Obras Poéticas de Gregório de Matos Guerra*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/zwJ01s>>. Acesso em: 15 dez. 2016.
- 23 OLIVEIRA, A. de. *Poesias*. Rio de Janeiro: Garnier, 1912–1928.
- 24 SANTA RITA DURÃO, J. de. *Caramuru*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/d4TAq5>>. Acesso em: 15 dez. 2016.
- 25 SANTA RITA DURÃO, J. de. *Caramuru: Poema Épico do Descobrimento da Bahia*. Lisboa: Régia Oficina Tipográfica, 1781. Disponível em: <<https://goo.gl/veDqS7>>. Acesso em: 15 dez. 2016.

- 26 SANTA RITA DURÃO, J. de. *Caramuru: Poema Épico do Descobrimento da Bahia*. Rio de Janeiro: Mamiliano da C. Honorato, 1878. Disponível em: <<https://goo.gl/yCve99>>. Acesso em: 15 dez. 2016.
- 27 SILVEIRA DE SOUSA, D. *Cancioneiro*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/LCwGc0>>. Acesso em: 15 dez. 2016.
- 28 SILVEIRA DE SOUSA, D. *Indeléveis*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/R8PMQF>>. Acesso em: 15 dez. 2016.
- 29 SILVEIRA DE SOUSA, D. *Lises e Martírios*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/ntTL1i>>. Acesso em: 15 dez. 2016.
- 30 SILVEIRA DE SOUSA, D. *Passos Dolorosos*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/YFcTxR>>. Acesso em: 15 dez. 2016.
- 31 SILVEIRA DE SOUSA, D. *Poemas Dispersos*. Florianópolis: Literatura Digital, 2016. Disponível em: <<https://goo.gl/RdeAMI>>. Acesso em: 15 dez. 2016.
- 32 SILVEIRA, D. *Obra Completa*. Organização: Lauro Junkes. Florianópolis: Academia Catarinense de Letras, 2009.

## Referências

- 33 ADAMS, L. D.; BIRNBAUM, D. J. Perspectives on Computer Programming for the Humanities. *Text Technology*, v. 7, n. 1, p. 1–17, 1997.
- 34 AGIRREZABAL, M.; ARRIETA, B.; ASTIGARRAGA, A.; HULDEN, M. ZeusScansion: a Tool for Scansion of English Poetry. In: 11TH INTERNATIONAL CONFERENCE ON FINITE STATE METHODS AND NATURAL LANGUAGE PROCESSING, 2013, p. 18–24.
- 35 ALMENOVA, A. B. *Разработка и реализация алгоритма анализа метрических характеристик поэтических текстов*. 2013. Diss. (Mestrado) – Новосибирский государственный университет. Autores em russo: А. Б. Альменова.

- 36 ALMUHAREB, A.; ALKHARASHI, I.; SAUD, L. A. L.; ALTUWAIJRI, H. Recognition of Classical Arabic Poems. In: 2ND WORKSHOP ON COMPUTATIONAL LINGUISTICS FOR LITERATURE, 2013, p. 9–16.
- 37 ALNAGDAWI, M. A.; RASHIDEH, H.; ABURUMMAN, A. F. Finding Arabic Poem Meter using Context Free Grammar. *Journal of Communications and Computer Engineering*, v. 3, n. 1, p. 52–59, 2013.
- 38 AMARAL, R. B.; CARVALHO, P. M.; CASEIRO, D. A.; TRANCOSO, I. M.; OLIVEIRA, L. C. Anotação fonética automática de corpora de fala transcritos ortograficamente. In: 4TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL PROCESSING OF PORTUGUESE, 1999.
- 39 ARAÚJO, P. A.; MAMEDE, N. J. Classificador de Poemas. In: CONFERÊNCIA CIENTÍFICA E TECNOLÓGICA EM ENGENHARIA, 2002.
- 40 ARAÚJO, P. A. M. *Classificação de Poemas e Sugestão das Palavras Finais dos Versos*. 2004. Diss. (Mestrado) – Universidade Técnica de Lisboa.
- 41 AYECH, H. E.; MAHFOUF, A.; ZRIBI, A. Reconnaissance de la métrique des poèmes arabes par les réseaux de neurones artificiels. In: 13ÈME CONFÉRENCE SUR LE TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES, 2006, p. 462–472.
- 42 BANDEIRA, M. *Itinerário de Pasárgada*. 4ª Edição. Rio de Janeiro: Nova Fronteira, 1984.
- 43 BARAKHNIN, V. B.; KOZHEMIKINA, O. I. Об автоматизации комплексного анализа русского поэтического текста. In: XIV всероссийская научная конференция электронные библиотеки: перспективные методы и технологии, электронные коллекции, 2012. Autores em russo: В. Б. Баракнин; О. Ю. Кожемякина.
- 44 BARAKHNIN, V. B.; KOZHEMIKINA, O. I.; ZABAİKIN, A. V. Алгоритмы комплексного анализа русских поэтических текстов с целью автоматизации процесса создания метрических справочников и конкордансов. In: XVII международная конференция аналитика и управление данными в областях с интенсивным использованием данных, 2015, p. 138–143. Autores em russo: В. Б. Баракнин; О. Ю. Кожемякина; А. В. Забайкин.

- 45 BARAKHNIN, V. B.; KOZHEMIKINA, O. I.; ZABAIIIN, A. V.; KHAIIATOVA, V. D. Автоматизация комплексного анализа русского поэтического текста: модели и алгоритмы. *Вестник НГУ. Серия: Информационные технологии*, v. 13, n. 3, p. 5–18, 2015. Autores em russo: В. Б. Барахнин; О. Ю. Кожемякина; А. В. Забайкин; В. Д. Хаятова.
- 46 BARBER, C.; BARBER, N. The Versification of *The Canterbury Tales*: A Computer-based Statistical Study. Part I. *Leeds Studies in English*, v. 21, p. 81–103, 1990.
- 47 BARBER, C.; BARBER, N. The Versification of *The Canterbury Tales*: A Computer-based Statistical Study. Part II. *Leeds Studies in English*, v. 22, p. 57–83, 1991.
- 48 BARQUIST, C. R.; SHIE, D. L. Computer Analysis of Alliteration in *Beowulf* Using Distinctive Feature Theory. *Literary and Linguistic Computing*, v. 6, n. 4, p. 274–280, 1991.
- 49 BARROS, M. J.; WEISS, C. Maximum Entropy Motivated Grapheme-to-Phoneme, Stress and Syllable Boundary Prediction for Portuguese Text-to-Speech. In: IV JORNADAS EN TECNOLOGÍA DEL HABLA, 2006, p. 177–182.
- 50 BEAUDOUIN, V. Mètre en règles. *Revue française de linguistique appliquée*, v. 9, p. 119–137, 2004.
- 51 BEAUDOUIN, V. *Rythme et rime de l'alexandrin classique: Étude empirique des 80 000 vers du théâtre de Corneille et Racine*. 2000. Tese (Doutorado) – École des Hautes Études en Sciences Sociales.
- 52 BEAUDOUIN, V. Rythme et univers lexicaux chez Corneille et Racine. In: 4ÈMES JOURNÉES INTERNATIONALES D'ANALYSE STATISTIQUE DES DONNÉES TEXTUELLES, 1998, p. 85–93.
- 53 BEAUDOUIN, V.; YVON, F. Contribution de la métrique à la stylométrie. In: 7ÈMES JOURNÉES INTERNATIONALES D'ANALYSE STATISTIQUE DES DONNÉES TEXTUELLES, 2004, p. 107–118.
- 54 BEAUDOUIN, V.; YVON, F. The Metrometer: a Tool for Analysing French Verse. *Literary and Linguistic Computing*, v. 11, n. 1, p. 23–31, 1996.
- 55 BERNET, C. Hasards de la rime. In: 7ÈMES JOURNÉES INTERNATIONALES D'ANALYSE STATISTIQUE DES DONNÉES TEXTUELLES, 2004, p. 148–159.

- 56 BERRY, D. M. Introduction: Understanding the Digital Humanities. In: BERRY, D. M. (Ed.). *Understanding Digital Humanities*. Basingstoke: Palgrave Macmillan, 2012. p. 1–20.
- 57 BERSCHIN, H. Automatische metrische Analyse lateinischer Verse. *Linguistik und Didaktik*, v. 1, p. 72–80, 1970.
- 58 BILAC, O.; PASSOS, G. *Tratado de Versificação*. Rio de Janeiro: Francisco Alves, 1905.
- 59 BIRNBAUM, D. J.; THORSEN, E. Markup and Meter: Using XML Tools to Teach a Computer to Think About Versification. In: BALISAGE: THE MARKUP CONFERENCE, 2015.
- 60 BISOL, L. O sândi e a ressilabação. *Letras de Hoje*, v. 31, n. 2, 1996.
- 61 BLUMENSTEIN, J.; DEUFERT, M.; GAERTNER, J. F. Elektronische Analyse der plautinischen Sprechverse: Ein Werkstattbericht. *eAQUA Working Papers*, v. 1, p. 101–107, 2010.
- 62 BOBENHAUSEN, K. Automatisches Metrisches Markup. In: DIGITAL HUMANITIES, 2009, p. 69–72.
- 63 BOBENHAUSEN, K. Automatisches metrisches Markup deutschsprachiger Gedichte. *Jahrbuch für Computerphilologie*, v. 9, p. 63–68, 2009.
- 64 BOBENHAUSEN, K.; HAMMERICH, B. Métrique littéraire, métrique linguistique et métrique algorithmique de l'allemand mises en jeu dans le programme Metricalizer<sup>2</sup>. *Langages*, v. 3, n. 199, p. 67–88, 2015.
- 65 БОЙКОВ, В. Н.; КАРИАЕВА, М. С.; СОКОЛОВ, В. А.; ПИЛЬЩИКОВ, И. А. Об автоматической спецификации стиха в информационно-аналитической системе. In: XVII международная конференция аналитика и управление данными в областях с интенсивным использованием данных, 2015, p. 144–151. Autores em russo: В. Н. Бойков; М. С. Каряева; В. А. Соколов; И. А. Пильщиков.
- 66 BRAGA, D.; RESENDE JR, F. G. V. Módulos de processamento de texto baseados em regras para sistemas de conversão Texto-Fala em PE. In: XXII ENCONTRO NACIONAL DA ASSOCIAÇÃO PORTUGUESA DE LINGUÍSTICA, 2007, p. 141–155.

- 67 BRASIL. Decreto nº 6.583, de 29 de setembro de 2008. Promulga o Acordo Ortográfico da Língua Portuguesa, assinado em Lisboa, em 16 de dezembro de 1990, 2008.
- 68 BREIDO, E. M. *Автоматический анализ метрики русского стиха*. 1996. Diss. (Mestrado) – Институт Русского Языка им. В. В. Виноградова. Autores em russo: E. M. Брейдо.
- 69 BRITTO, P. H. Uma Experiência de Autotradução. *Philia & Filia*, v. 2, n. 1, 2013.
- 70 CAILLET, M.; ROISIN, C.; CARRIVE, J. Multimedia Applications for Playing with Digitized Theater Performances. *Multimedia Tools and Applications*, v. 73, n. 3, p. 1777–1793, 2014.
- 71 CARVALHO, R. S. *Identificação de Padrões de Versificação em Texto em Prosa da Língua Portuguesa*. Universidade Estadual de Feira de Santana, 2015.
- 72 CHISHLOM, D. Phonology and Style: A Computer-Assisted Approach to German Verse. *Computers and the Humanities*, v. 15, p. 199–210, 1981.
- 73 CHOCIAY, R. A Identidade Formal do Decassílabo em “O Uruguai”. *Revista de Letras*, v. 34, p. 229–243, 1994.
- 74 CHOCIAY, R. *Os Metros do Boca: Teoria do Verso em Gregório de Matos*. São Paulo: UNESP, 1993.
- 75 CHOCIAY, R. *Teoria do Verso*. São Paulo: McGraw-Hill do Brasil, 1974.
- 76 CIOBANU, A. M.; DINU, L. P. On the Romanian Rhyme Detection. In: 24TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 2012, p. 87–93.
- 77 COHEN, R. *500 Cantigas d'Amigo*. Porto: Campo das Letras, 2003.
- 78 COUTO, I.; NETO, N.; TADAIESKY, V.; KLAUTAU, A.; MAIA, R. An Open Source HMM-based Text-to-Speech System for Brazilian Portuguese. In: INTERNATIONAL TELECOMMUNICATIONS SYMPOSIUM, 2010.
- 79 DELENTE, É.; RENAULT, R. Annotation automatique de textes versifiés. *Schedae*, p. 39–52, 2011.
- 80 DELENTE, É.; RENAULT, R. Outils et métrique : Un tour d'horizon. *Langages*, v. 3, n. 199, p. 5–22, 2015.

- 81 DELENTE, É.; RENAULT, R. Projet Anamètre : Le calcul du mètre des vers complexes. *Langages*, v. 3, n. 199, p. 125–148, 2015.
- 82 DELENTE, É.; RENAULT, R. Traitement automatique des formes métriques des textes versifiés. In: 22ÈME CONFÉRENCE SUR LE TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES, 2015, p. 432–438.
- 83 DELMONTE, R. A Computational Approach to Poetic Structure, Rhythm and Rhyme. In: 1ST ITALIAN CONFERENCE ON COMPUTATIONAL LINGUISTICS, 2014, p. 144–150.
- 84 DELMONTE, R. Visualizing Poetry with SPARSAR – Visual Maps from Poetic Content. In: 4TH WORKSHOP ON COMPUTATIONAL LINGUISTICS FOR LITERATURE, 2015, p. 68–78.
- 85 DEUFERT, M.; BLUMENSTEIN, J.; TREBESIU, A.; BEYER, S.; BÜCHLER, M. Objective Detection of Plautus' Rules by Computer Support. In: DIGITAL HUMANITIES 2010, 2010, p. 126–127.
- 86 DONOW, H. S. Prosody and the Computer: A Text Processor for Stylistic Analysis. In: SPRING JOINT COMPUTER CONFERENCE, 1970, p. 287–295.
- 87 DUQUE-ESTRADA, O. *A Arte de Fazer Versos*. 2. ed. Rio de Janeiro: Livraria Francisco Alves, 1914.
- 88 EGASHIRA, F. *Síntese de Voz a Partir de Texto para a Língua Portuguesa*. 1992. Diss. (Mestrado) – Universidade Estadual de Campinas.
- 89 ESTES, A.; HENCH, C. Supervised Machine Learning for Hybrid Meter. In: 5TH WORKSHOP ON COMPUTATIONAL LINGUISTICS FOR LITERATURE, 2016, p. 1–8.
- 90 ÉVRARD, É. Scansion automatique de l'hexamètre grec. *Revue de l'Organisation Internationale pour l'Etude des Langues Anciennes par Ordinateur*, v. 8, n. 4, p. 1–33, 1972.
- 91 FOLEY, J. M. A Computer Analysis of Metrical Patterns in *Beowulf*. *Computers and the Humanities*, v. 12, p. 71–80, 1978.
- 92 FUSI, D. A Multilanguage, Modular Framework for Metrical Analysis: IT Patterns and Theoretical Issues. *Langages*, v. 3, n. 199, p. 41–66, 2015.
- 93 FUSI, D. An Expert System for the Classical Languages: Metrical Analysis Components. *Lexis*, v. 27, p. 25–45, 2008.



- 94 FUSI, D. Appunti sulla prosodia del Lussorio di Shackleton-Bailey: alcune questioni di metodo. In: BERTINI, F. (Ed.). *Luxoriana*. Genova: Darficlet, 2002. p. 629–630.
- 95 FUSI, D. Fra metrica e linguistica: per la contestualizzazione di alcune leggi esametriche. In: DI LORENZO, E. (Ed.). *Lesametro greco e latino: analisi, problemi e prospettive*. Napoli: Guida, 2002. p. 33–63.
- 96 GASPAROV, M. L. *Очерк истории европейского стиха*. Москва: Фортуна Лимитед, 2003. Autores em russo: М. Л. Гаспаров.
- 97 GENZEL, D.; USZKOREIT, J.; OCH, F. “Poetic” Statistical Machine Translation: Rhyme and Meter. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2010, p. 158–166.
- 98 GERVÁS, P. A Logic Programming Application for the Analysis of Spanish Verse. In: 1ST INTERNATIONAL CONFERENCE ON COMPUTATIONAL LOGIC, 2000, p. 1330–1344.
- 99 GOLDSTEIN, N. S. *Versos, sons, ritmos*. 14. ed. São Paulo: Ática, 2008.
- 100 GOMES, L. d. C. T. *Sistema de Conversão Texto-Fala para a Língua Portuguesa Utilizando a Abordagem de Síntese por Regras*. 1998. Diss. (Mestrado) – Universidade Estadual de Campinas.
- 101 GOUVEIA, P. D. F.; TEIXEIRA, J. P. R.; SILVA FREITAS, D. R. da. Divisão Silábica Automática do Texto Escrito e Falado. In: 5TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL PROCESSING OF PORTUGUESE, 2000.
- 102 GREENBERG, N. A. Scansion purement automatique de l’hexamètre dactylique. *Revue de l’Organisation Internationale pour l’Etude des Langues Anciennes par Ordinateur*, v. 3, n. 3, p. 1–30, 1967.
- 103 GREENE, E.; BODRUMLU, T.; KNIGHT, K. Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation. In: 2010 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2010, p. 524–533.
- 104 HANSEN, J. A. Autorialia, obra e público na poesia colonial luso-brasileira atribuída a Gregório de Matos e Guerra. *Ellipsis*, v. 12, p. 91–117, 2014.
- 105 HARTMAN, C. *The Scandroid*. New London, 2005. Disponível em: <<http://oak.conncoll.edu/cohar/Programs.htm>>.

- 106 HAYLES, N. K. How We Think: Transforming Power and Digital Technologies. In: BERRY, D. M. (Ed.). *Understanding Digital Humanities*. Basingstoke: Palgrave Macmillan, 2012. p. 42–66.
- 107 HAYWARD, M. A Connectionist Model of Poetic Meter. *Poetics*, v. 20, p. 303–317, 1991.
- 108 HAYWARD, M. Analysis of a Corpus of Poetry by a Connectionist Model of Poetic Meter. *Poetics*, v. 24, p. 1–11, 1996.
- 109 HERVÉ, N. Élaboration et utilisation des relevés métriques. Réflexions théoriques sur un cas pratique : Le relevé Clément Marot. *Langages*, v. 3, n. 199, p. 23–40, 2015.
- 110 HIDLEY, G. R. Some Thoughts Concerning the Application of Software Tools in Support of Old English Poetic Studies. *Literary and Linguistic Computing*, v. 1, n. 3, p. 156–162, 1986.
- 111 HIRJEE, H. *Rhyme, Rhythm, and Rhubarb: Using Probabilistic Methods to Analyze Hip Hop, Poetry, and Misheard Lyrics*. 2010. Diss. (Mestrado) – University of Waterloo.
- 112 HIRJEE, H.; BROWN, D. G. Automatic Detection of Internal and Imperfect Rhymes in Rap Lyrics. In: 10TH INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 2009, p. 711–716.
- 113 HIRJEE, H.; BROWN, D. G. Using Automated Rhyme Detection to Characterize Rhyming Style in Rap Music. *Empirical Musicology Review*, v. 5, n. 4, p. 121–145, 2010.
- 114 IBRAHIM, R.; PLECHÁČ, P. Towards the Automatic Analysis of Czech Verse. In: FORMAL Methods in Poetics. Lüdenscheid: RAM-Verlag, 2011. p. 295–305.
- 115 JIMÉNEZ, F. M. y. Contribución al Análisis de la Estructura del Hexámetro Latino Mediante Ordenadores Electrónicos. *Revista Española de Lingüística*, v. 4, n. 1, p. 230–237, 1974.
- 116 JONES, F. P. A Binary-Octal Code for Analyzing Hexameters. *Transactions and Proceedings of the American Philological Association*, v. 97, p. 275–280, 1966.
- 117 KAVANAGH, F. *Analysis of a Phonetic and Rule Based Algorithm Approach to Determine Rhyme Categories and Patterns in Verse*. 2008. Diss. (Mestrado) – Open University.

- 118 KOULOUGHLI, D. E. Traitement automatique de la métrique arabe : réalisations et perspectives. *Bulletin d'études orientales*, v. LIX, p. 17–31, 2010.
- 119 KOZMIN, A. V. Автоматический анализ стиха в системе Starling. In: диалог 2016, 2006, p. 265–268. Autores em russo: A. B. Козьмин.
- 120 KURT, A.; KARA, M. An Algorithm for the Detection and Analysis of Arud Meter in Diwan Poetry. *Turkish Journal of Electrical Engineering & Computer Sciences*, v. 20, n. 6, p. 948–963, 2012.
- 121 LATSCH, V. L. *Desenvolvimento de um Sistema de Conversão Texto-Fala com Modelagem de Prosódia*. 2011. Tese (Doutorado) – Universidade Federal do Rio de Janeiro.
- 122 MAMEDE, N.; TRANCOSO, I.; ARAÚJO, P.; VIANA, C. An Electronic Assistant for Poetry Writing. In: 9TH IBERO-AMERICAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2004, p. 286–294.
- 123 MAMEDE, N.; TRANCOSO, I.; ARAÚJO, P.; VIANA, C. Poetry Assistant. In: 8TH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 2004.
- 124 MARQUES, J. A. D. *Sistema de Apoio à Escrita de Poemas*. 2008. Diss. (Mestrado) – Universidade Técnica de Lisboa.
- 125 MATTOSO, G. *Tractado de Orthographia Lusophona*. [S.l.]: Elson Fróes, 2011. Disponível em: <<https://goo.gl/Y2IwKy>>.
- 126 MATTOSO, G. *Tratado de Versificação*. São Paulo: Annablume, 2010.
- 127 MAYRHOFFER, C. M. Scansion and Analysis of Prakrit Verses by Text-Processing Programs. *Revue informatique et statistique dans les sciences humaines*, v. XXIII, p. 99–110, 1987.
- 128 MCCURDY, N.; LEIN, J.; COLES, K.; MEYER, M. Poemage: Visualizing the Sonic Topology of a Poem. *IEEE Transactions on Visualization and Computer Graphics*, v. 22, n. 1, p. 439–448, 2016.
- 129 MCCURDY, N.; SRIKUMAR, V.; MEYER, M. RhymeDesign: A Tool for Analyzing Sonic Devices in Poetry. In: 4TH WORKSHOP ON COMPUTATIONAL LINGUISTICS FOR LITERATURE, 2015, p. 12–22.
- 130 MITTMANN, A.; WANGENHEIM, A. von; SANTOS, A. L. dos. A Multi-Level Visualization Scheme for Poetry. In: 20TH INTERNATIONAL CONFERENCE ON INFORMATION VISUALISATION, 2016.

- 131 MITTMANN, A.; WANGENHEIM, A. von; SANTOS, A. L. dos. Aoidos: A System for the Automatic Scansion of Poetry Written in Portuguese. In: 17TH INTERNATIONAL CONFERENCE ON INTELLIGENT TEXT PROCESSING AND COMPUTATIONAL LINGUISTICS, 2016.
- 132 MOISÉS, C. F. Indagações sobre o verso livre. *Dicta & Contradicta*, n. 4, 2009.
- 133 MORENO, J. L. Un Método para el Tratamiento Informático de Materiales Latinos en Verso. *Emerita*, v. 55, n. 1, p. 15–30, 1987.
- 134 MORETTI, F. *Distant reading*. London: Verso, 2013.
- 135 NAVARRO-COLORADO, B. A Computational Linguistic Approach to Spanish Golden Age Sonnets: Metrical and Semantic Aspects. In: 4TH WORKSHOP ON COMPUTATIONAL LINGUISTICS FOR LITERATURE, 2015, p. 105–113.
- 136 NAVARRO-COLORADO, B.; LAFOZ, M. R.; SÁNCHEZ, N. Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation. In: 9TH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2016.
- 137 NAVARRO-COLORADO, B.; LAFOZ, M. R.; TRIGUEROS, S. J.; SÁNCHEZ, N. Compilación y Anotación Métrica de un Corpus de Sonetos del Siglo de Oro. In: II CONGRESO INTERNACIONAL HUMANIDADES DIGITALES HISPÁNICAS, 2015.
- 138 NETO, N.; ROCHA, W.; SOUSA, G. An open-source rule-based syllabification tool for Brazilian Portuguese. *Journal of the Brazilian Computer Society*, v. 21, n. 1, p. 1–10, 2015.
- 139 OLIVEIRA, L. C.; VIANA, M. C.; TRANCOSO, I. M. A Rule-Based Text-to-Speech System for Portuguese. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1992. v. 2, p. 73–76.
- 140 OLIVEIRA, L. C.; VIANA, M. C.; TRANCOSO, I. M. DIXI — Portuguese Text-to-Speech System. In: 2ND EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, 1991, p. 1239–1242.
- 141 OPARA, K. R. Grammatical Rhymes in Polish Poetry: A Quantitative Analysis. *Digital Scholarship in the Humanities*, v. 30, n. 4, p. 589–598, 2015.

- 142 OTT, W. Metrical Analysis of Latin Hexameter by Computer. *Revue de l'Organisation Internationale pour l'Etude des Langues Anciennes par Ordinateur*, v. 2, n. 4, p. 7–23, 1966.
- 143 OTT, W. Metrical Analysis of Latin Hexameter by Computer - 2d part. *Revue de l'Organisation Internationale pour l'Etude des Langues Anciennes par Ordinateur*, v. 3, n. 1, p. 39–96, 1967.
- 144 OTT, W. *Metrische Analysen zur Ars Poetica des Horaz*. Goppingen: Alfred Kummerle, 1970.
- 145 PAPAITSOS, E. C. Computerized Scansion of Ancient Greek Hexameter. *Literary and Linguistic Computing*, v. 26, n. 1, p. 57–69, 2010.
- 146 PILSHCHIKOV, I. A.; STAROSTIN, A. S. Автоматическое распознавание метра: проблемы и решения. In: славянский стих IX, 2008, p. 492–498. Autores em russo: И. А. Пильщиков; А. С. Старостин.
- 147 PILSHCHIKOV, I. A.; STAROSTIN, A. S. Основные проблемы автоматизации базовых процедур ритмико-синтаксического анализа силлабо-тонических текстов. In: Национальный корпус русского языка: 2006 – 2008. Новые результаты и перспективы. Санкт-Петербург: Нестор-История, 2009. Autores em russo: И. А. Пильщиков, А. С. Старостин.
- 148 PILSHCHIKOV, I.; STAROSTIN, A. Automated Analysis of Poetic Texts and the Problem of Verse Meter. In: KÜPER, C. (Ed.). *Current Trends in Metrical Analysis*. Bern: Peter Lang, 2011. p. 133–140.
- 149 PILSHCHIKOV, I.; STAROSTIN, A. Reconnaissance automatique des mètres des vers russes : Une approche statistique sur corpus. *Langages*, v. 3, n. 199, p. 89–106, 2015.
- 150 PLAMONDON, M. R. Computer-Assisted Phonetic Analysis of English Poetry: A Preliminary Case Study of Browning and Tennyson. *Text Technology*, v. 14, n. 2, p. 153–175, 2005.
- 151 PLAMONDON, M. R. Virtual Verse Analysis: Analysing Patterns in Poetry. *Literary and Linguistic Computing*, v. 21, p. 127–141, Suppl. 1 2006.
- 152 PLECHÁČ, P. Czech Verse Processing System KVĚTA — Phonetic and Metrical Components. *Glottology*, v. 7, 2 2016.

- 153 PLECHÁČ, P.; KOLÁR, R. The Corpus of Czech Verse. *Studia Metrica et Poetica*, v. 2, n. 1, 2015.
- 154 PROENÇA, M. C. *Ritmo e Poesia*. 1. ed. Rio de Janeiro: Organização Simões, 1955.
- 155 RAINSFORD, T. M.; SCRIVNER, O. Metrical Annotation for a Verse Treebank. In: 13TH INTERNATIONAL WORKSHOP ON TREEBANKS AND LINGUISTIC THEORIES, 2014, p. 149–159.
- 156 RAKSHIT, G.; GHOSH, A.; BHATTACHARYA, P.; HAFFARI, G. Automated Analysis of Bangla Poetry for Classification and Poet Identification. In: 12TH INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING, 2015.
- 157 RAMA, N.; LAKSHMANAN, M. A Computational Algorithm for Metrical Classification of Verse. *International Journal of Computer Science Issues*, v. 7, n. 2, p. 46–53, 2010.
- 158 REDDY, S.; KNIGHT, K. Unsupervised Discovery of Rhyme Schemes. In: 49TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 2011, p. 77–82.
- 159 ROBey, D. Scanning Dante's the *Divine Comedy*. A Computer-based Approach. *Literary and Linguistic Computing*, v. 8, n. 2, p. 81–84, 1993.
- 160 ROBINSON, J. R. *Colors of Poetry: Computational Deconstruction*. 1998. Diss. (Mestrado) – Georgia State University.
- 161 ROCHA, J. M. N. da. *Como Construir um Leitor de Poesia*. 2008. Tese (Doutorado) – Universidade de Lisboa.
- 162 ROUBAUD, J. DYNASTIE : études sur le vers français, sur l'alexandrin classique. *Cahiers de poétique comparée*, n. 13, p. 47–109, 1986. Première partie.
- 163 ROUBAUD, J. DYNASTIE : études sur le vers français, sur l'alexandrin classique. *Cahiers de poétique comparée*, n. 16, p. 41–60, 1988. Deuxième partie.
- 164 SGALLOVÁ, K. Thesaurus českých meter. *Česká literatura*, v. 47, n. 3, p. 286–289, 1999.
- 165 SGALLOVÁ, K. Využití moderní techniky při rozboru verše. *Česká literatura*, v. 12, n. 2, p. 158–168, 1964.
- 166 TAVARES, H. Ú. d. C. *Teoria Literária*. 4. ed. Belo Horizonte: Bernardo Álvares S.A., 1969.

- 167 TEI CONSORTIUM. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Versão 3.0.0. [S.l.], 2016. Disponível em: <<http://www.tei-c.org/Guidelines/P5/>>.
- 168 TEIXEIRA, J. P. R.; SILVA FREITAS, D. R. da; GOUVEIA, P. D. F.; OLASZY, G.; NÉMETH, G. MULTIVOX — Conversor Texto Fala Para Português. In: 3RD INTERNATIONAL CONFERENCE ON COMPUTATIONAL PROCESSING OF PORTUGUESE, 1998.
- 169 TIZHOOSH, H. R.; DARA, R. A. On Poem Recognition. *Pattern Analysis and Applications*, v. 9, p. 325–338, 2006.
- 170 WILKENS, M. Canons, Close Reading, and the Evolution of Method. In: GOLD, M. K. (Ed.). *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. p. 249–258.
- 171 WUJASTYK, D. Automatic Scansion of Sanskrit Poetry for Authorship Criteria. *Association for Literary and Linguistic Computing Bulletin*, v. 6, n. 2, p. 122–135, 1978.
- 172 Y. OUSAKA M. Yamazaki, M. M. Automatic Analysis of the Canon in Middle Indo-Aryan by Personal Computer. *Literary and Linguistic Computing*, v. 9, n. 2, p. 125–136, 1994.
- 173 Y. OUSAKA, M. Y. Automatic Analysis of the Canon in Middle Indo-Aryan by Personal Computer II. *Literary and Linguistic Computing*, v. 11, n. 1, p. 9–17, 1996.





## A. VERSOS INESCANDÍVEIS

Este apêndice lista todos os versos que foram considerados inescandíveis durante as revisões dos 10 *corpora* poéticos principais. O número entre o código e o verso é o tamanho silábico esperado. Alguns versos inescandíveis do *corpus* FIG encontram-se em poemas polimétricos; neste caso, estão marcados com  $\leq 12$ , pois não se sabe o número exato de sílabas que eles deveriam possuir.

### ***Corpus* CAM**

Nenhum verso inescandível neste *corpus*.

### ***Corpus* MAT**

MAT:1.1.1.1.2	10	Que nos quer governar a cabana, e vinha,
MAT:1.5.1.1.16	7	luz, sol, mortos, e vivos:
MAT:2.1.7.1	7	De Jesus tanto agrado leva
MAT:2.22.6.5	7	dá-lhes as contas do Rosário.
MAT:4.27.1.4	10	E faz servir seu cu de cocó=.
MAT:4.33.1.20	7	ter-lhes ódio, isso fora ruim.
MAT:5.8.1.1	10	Bem-vindo seja, Senhor, Vossa ilustríssima
MAT:5.14.2.8	7	já filósofo, ou já letrado,
MAT:5.29.5.6	7	Se este Médico, este tramposo
MAT:5.35.4.1	10	Enfim Papagaio humano te perdeste,
MAT:5.48.7.1	10	Não te desvaneça andar-te a puta ao rabo,
MAT:5.48.9.4	6	E se há puta, que te atura
MAT:7.1.13.5	10	Não sofrem dilatação os da Bahia
MAT:7.1.16.6	10	Última resolução da cobardia:
MAT:7.2.1.3	10	Por mostrar, se da morte da vida é choro,
MAT:7.9.2.2	7	e alegrou-se tanto o monete,
MAT:8.2.1.15	7	se tem força terá a verdade.
MAT:9.10.80.6	7	vendo, que lha surraram
MAT:10.15.3.3	10	Nasceu flor, a quem um sol faz tanto dano?
MAT:10.23.4.1	10	Chora festivo já, ó cristal sonoro,

MAT:12.3.9.1	7	Olá, ô! chegou o Tudesco:
MAT:12.8.1.37	7	disse eu, e a vocês lhes esquece:
MAT:12.14.16.3	7	qual o Moço de Dom Quixote,
MAT:15.9.7.10	7	e o do Padre Manuel Abrás.
MAT:15.12.3.2	10	Que esse lhe dá de render o pão da me-,
MAT:18.4.2.9	10	Não entreis em Aganipe mais na barca,
MAT:19.3.1.19	7	como garbo de mais da marca,
MAT:22.17.2.2	10	Mas ei-lo vai, tudo o que é de amor, obriga,
MAT:24.20.2.7	7	que no meu coração não erra
MAT:27.1.1.109	7	Fretam-nos finalmente
MAT:29.3.3.10	7	Senhora Lima? Lim. Que hei de fazer?
MAT:29.21.3.3	10	Até nascer-lhe de Pero o sol formoso:
MAT:29.22.7.8	7	dou ao demo o quies vel qui,
MAT:29.23.10.4	10	ao peito espaldar debaixo da viseira.
MAT:29.47.5.2	5	sabia todo o mundo,
MAT:29.50.2.3	10	Lembra-te daquele amigo, a quem deixaste
MAT:29.71.16.2	7	e que se ao som do seu desvelo
MAT:29.102.1.26	7	se já verdade em pataratas,
MAT:29.122.1.47	7	mas vós tendes este gostinho,
MAT:33.4.4.8	7	cento, e cinqüenta réis lhe achais,
MAT:35.6.1.58	7	quando me disse “ecce”
MAT:36.4.1.97	7	Terra, que não parece
MAT:38.18.1.4	10	Fatal presunção do sofrimento!
MAT:38.21.1.41	7	É tal Floral, não sei, se o diga,
MAT:39.7.1.45	7	“Tiro lico tico, ré fá.”

### ***Corpus DUR***

Nenhum verso inescandível neste *corpus*.

### ***Corpus COS***

Nenhum verso inescandível neste *corpus*.

### ***Corpus GON***

Nenhum verso inescandível neste *corpus*.

## **Corpus MAG**

Nenhum verso inescandível neste *corpus*.

## **Corpus PIN**

PIN.27.25.3	10	Mas raposa, por astúcia abalizada.
PIN.58.46.2	10	Aos olhos, dera em fusão, no forno ardente,
PIN.71.23.2	10	Pareça injusta é, de péssima heresia,
PIN.78.14.2	10	O que de outros aos louvores mais se estende:
PIN.78.36.1	10	"Na dura penha, que se interpõe ao leito
PIN.93.16.3	10	O amor teu, e mais que tudo a Deus destina.
PIN.94.11.1	10	Qual a dama, que à virtude cultos vende
PIN.100.29.1	10	E lá na profundidade vi que se interna

## **Corpus ANJ**

ANJ.261.1.1	10	Eurípedes — o maior fisiologista
-------------	----	----------------------------------

## **Corpus FIG**

FIG:1.4.2.2	12	De perfumes e canção. Mas, o mar, num repente,
FIG:1.19.2.2	12	Por acaso andarei, neste mundo, a ambicionada
FIG:1.22.2.2	12	Pelas praias e campos, existe um burburinho
FIG:1.28.2.2	12	Senão perto de ti, com seus matizes?
FIG:1.39.1.1	12	Aos meus braços encostas o teu peito, Maria,
FIG:1.68.2.3	12	O meu noivo vagara aflito, envolvido nas fráguas,
FIG:1.104.2.2	12	Dos nossos braços, eis-nos na pescaria...
FIG:1.115.4.2	12	Esses seios de praias, alvíssimos, de opala,
FIG:1.132.2.4	12	Uivam nas praias os cães e mugem bois, errantes...
FIG:2.1.17.1	7	la-lhes um sorriso brando
FIG:2.7.2.2	10	E, lépido passa, entre elas, soluçando...
FIG:2.14.13.1	≤ 12	E que saudades havia a toda hora do dia!
FIG:2.17.48.2	7	E se veem os filhos chorando,
FIG:2.23.31.2	7	Como dos chupins irrequietos:
FIG:2.24.29.2	7	Das praias da luz cheia.
FIG:2.25.5.15	≤ 12	Que, na terra, deixou de olhá-las e compreendê-las,
FIG:2.26.22.2	≤ 12	meiga vaquinha, que se encheram de eterno sentimento,
FIG:2.27.3.15	≤ 12	Sob a piedosa cruz, a misericordiosa cruz

FIG:4.39.1.2	10	E nos unimos apaixonadamente?
FIG:4.46.2.3	10	Transfigurando-me a vida atra e nefasta,
FIG:4.47.2.2	10	Em derredor de ti; cantando esse coro;
FIG:4.56.4.1	10	No teu amor, portanto, visto,
FIG:4.70.2.2	10	Da luz do afago, que se desfaz em mantos...
FIG:5.17.3.1	10	Afastai-vos de mim... Mas de que jeito, e como
FIG:5.20.1.3	10	Que os corações por mais austeros ao vê-los,
FIG:5.23.1.2	10	Que dá ranger de dentes e convulsões,
FIG:5.24.1.3	10	Unta-lhe as botas prateadas um vassalo,
FIG:5.43.3.2	10	Que têm todos os fulgíssimos encantos
FIG:5.45.1.2	10	São assim tão frios e tão duros!
FIG:5.57.2.1	10	Trato de amá-lo, e trato de compreendê-lo,
FIG:5.64.2.2	10	Todas as forças — e que, nessa hora, delas
FIG:5.65.6.4	12	E há músicas no seu acariciante falar.
FIG:5.77.1.2	10	E as açucenas tornam-se celestiais...

### **Corpus SIL**

SIL:1.2.5.9	7	ao bater dos remos
SIL:1.33.1.6	6	nas ondas suspira.
SIL:1.57.2.5	10	ou, nesta desilusão amarga, infinda,
SIL:1.71.5.2	7	meiga Flor da Judeia,
SIL:1.88.1.5	10	O preito de gratidão, vinde, singelos,
SIL:1.104.1.1	10	Amo-te, mar, oh! quando em tuas águas quedas
SIL:1.112.1.1	12	Cantar a Natureza, de Deus o Nome Santo
SIL:1.112.3.9	12	cantar a Natureza, de Deus o Santo Nome
SIL:1.112.4.1	12	Cantar a Natureza, de Deus o Nome Santo
SIL:1.125.3.1	10	— Coração, alma, vida de penas,
SIL:1.128.1.3	10	rebrilhando mar em frisas d'ouro.
SIL:1.139.5.2	10	ao tronco protetor s'enlaça,
SIL:2.29.12.4	7	Tem na mão fechada,
SIL:2.37.1.9	9	Traz doçuras, traz os primores,
SIL:4.2.1.7	10	Eu, triste, pergunto ao sol, às brisas, ao silêncio
SIL:4.37.2.3	12	Ela do beijo letal da Morte adormecida,
SIL:5.25.1.6	7	ou quando for combate,
SIL:5.31.1.8	7	evitaria mil pesares,
SIL:5.61.1.4	7	as grandes datas do meu povo.
SIL:5.66.1.10	7	tenhas em mente a alheia dor".
SIL:5.107.1.2	10	que dentro de um coração terno palpita

SIL:5.107.1.3	10	tu gemes qual rola magoada,
SIL:5.112.1.6	7	Nascida no fundo do mar
SIL:5.118.2.4	10	depois que aurora o céu da noite dardeja.
SIL:5.125.6.7	4	traz-nos a bonança,
SIL:5.126.26.8	7	primeiro, aceso com piedade,
SIL:5.126.46.2	6	tão bem montado,
SIL:5.132.4.12	10	Às vezes, mar ofegante a s'espriar mansinho
SIL:5.140.8.1	7	A Virgem Maria junto aos seios
SIL:5.146.1.4	10	atravessei o mar, sombrio.
SIL:5.176.4.3	7	Veste solo gazil
SIL:5.180.2.3	10	Sim! Somos Livres! Infâmia, na verdade,
SIL:5.181.3.4	11	Cobrinha de flores a Terra Cruz.
SIL:5.187.2.3	10	Os pequenos pressurosos, ledos,
SIL:5.187.7.2	10	E respeitoso beija o símbolo sagrado;
SIL:5.202.3.5	10	Mas... Mas... o quê? Beatriz! Achaste
SIL:5.209.3.3	10	Inculto... Oh! vamos despertá-lo,
SIL:5.211.1.1	12	"Cruz imortal! Símbolo da Dor e Glória!
SIL:5.212.1.2	12	Que nos meus sonhos vens, como doce claridade,
SIL:5.233.1.2	10	Sem que minh'alma ocupe-se toda inteira,
SIL:5.236.2.4	12	Um Livro em que do talento as perlas foi guardando.
SIL:5.253.5.3	9	Dos rosários os lenitivos
SIL:5.261.3.1	9	— Amor — só às almas ensina



## B. REGRAS PROSÓDICAS

A partir da próxima página são listadas as 159 regras prosódicas do Aoidos, suficientes para a escansão dos mais de 100.000 versos do experimento principal. As regras estão ordenadas por camadas (de 1 a 10) e, dentro de cada camada, pelo número de aplicações no experimento principal, com as regras mais usadas primeiro. O número de identificação de cada regra é seguido da categoria à qual pertence (por exemplo, sinérese) e do número de aplicações por mil versos no experimento principal. As regras que estão entre as 20 que todos os 10 *corpora* utilizam são marcadas com uma estrela: ★. Cada regra possui um conjunto de parâmetros, que são passados à máquina de regras descrita no Capítulo 5. Por fim, um exemplo de aplicação é dado para cada regra, incluindo a escansão final do verso utilizado no exemplo.

## Camada 1

**Regra 905** acento 0,4‰

---

**ber:** /(kɪ)\$/ **msub:** “k’e” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **mpts:** 1 **mton:** (tornar tônica)

MAT:38.21.1.59 porque é tão divina, que

1 2 3 4 5 6 7 8  
*antes* pur|kɪ|’ɛ|’tāw|di|’vi|na|kɪ  
*depois* pur|kɪ|’ɛ|’tāw|di|’vi|na|’ke

1 2 3 4 5 6 7  
*final* pur|’kɛ|’tāw|di|’vi|na|’ke  
*final* por-|que é|tão|di-vi-na,|que

**Regra 904** acento 0,1‰

---

**ber:** /(pʊrkɪ)\$/ **msub:** “pʊrk’e” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona)

**blim:** (inicial, final) **mpts:** 1 **mton:** (nada, tornar tônica)

MAT:6.18.1.3 estimo muito: porque

1 2 3 4 5 6 7  
*antes* es|’ti|mu|’muj|tu|pur|kɪ  
*depois* es|’ti|mu|’muj|tu|pur|’ke

1 2 3 4 5 6 7  
*final* es|’ti|mu|’muj|tu|pur|’ke  
*final* es-|ti-mo-mui-to:|por-que

**Regra 907** acento 0,1‰

---

**ber:** /(ũ)\$/ **msub:** “ũ” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **mpts:** 1 **mton:** (tornar tônica)

MAT:24.20.3.9 pela minha vida um,

1 2 3 4 5 6 7  
*antes* ’pɛ|la|’mi|’na|’vi|da|ũ  
*depois* ’pɛ|la|’mi|’na|’vi|da|’ũ

1 2 3 4 5 6 7  
*final* ’pɛ|la|’mi|na|’vi|da|’ũ  
*final* pe-la-mi-nha-vi-da-um,



# Camada 1 (cont.)

**Regra 903** acento < 0,1‰

**ber:** /(uma)\$/ **msub:** “uma” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona)

**blim:** (inicial, final) **mpts:** 1 **mton:** (tornar tônica, nada)

ANJ.55.9.3 A conta recomeço, em ânsias: — Uma...

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	a	kō	ta	ve	ko	mē	su	ē	ā	sja	u	ma
<i>depois</i>	a	kō	ta	ve	ko	mē	su	ē	ā	sja	u	ma
<i>final</i>	a	kō	ta	ve	ko	mē	swē	ā	sja	zu	ma	
<i>final</i>	A	con	ta	re	co	me	ço	em	ân	sia	s	— U
												ma...

**Regra 902** acento < 0,1‰

**ber:** /(a)\$/ **msub:** “a” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **mpts:** 10 **mton:** (tornar tônica)

ANJ.217.1.20 Reduz os vegetais receptáculos a

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	ve	dus	us	ve	ze	tajs	ve	sep	ta	ku	lus	a
<i>depois</i>	ve	dus	us	ve	ze	tajs	ve	sep	ta	ku	lus	a
<i>final</i>	ve	du	zuz	ve	ze	tajz	ve	sep	ta	ku	lu	za
<i>final</i>	Re	du	z os	ve	ge	tais	re	cep	tá	cu	lo	s

**Regra 906** acento < 0,1‰

**ber:** /(si)\$/ **msub:** “s’e” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **mpts:** 10 **mton:** (tornar tônica)

MAT:15.5.1.3 Canoros pasmos te prevejo, se

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	ka	no	rus	paz	mus	ti	pre	ve	zu	si
<i>depois</i>	ka	no	rus	paz	mus	ti	pre	ve	zu	se
<i>final</i>	ka	no	rus	paz	mus	ti	pre	ve	zu	se
<i>final</i>	Ca	no	ros	pas	mos	te	pre	ve	jo	se

## Camada 1 (cont.)

**Regra 900**      acento      < 0,1‰

---

**ber:** /⟨nas⟩\$/ **msub:** “n’as” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **mpts:** 1 **mton:** (tornar tônica)

ANJ.54.2.2 Do céu, em reflexos, nas

	1	2	3	4	5	6	7
<i>antes</i>	du	ˈsew	ẽ	ˈʁe	ˈfle	ˈʃus	nas
<i>depois</i>	du	ˈsew	ẽ	ˈʁe	ˈfle	ˈʃus	nas
	1	2	3	4	5	6	7
<i>final</i>	du	ˈsew	ẽ	ˈʁe	ˈfle	ˈʃuz	nas
<i>final</i>	Do	ˈcéu,	em	re	ˈfle	ˈxos,	nas

**Regra 901**      acento      < 0,1‰

---

**ber:** /⟨pur⟩\$/ **msub:** “p’or” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **mpts:** 1 **mton:** (tornar tônica)

ANJ.54.3.2 Cai a luz dos astros por

	1	2	3	4	5	6	7
<i>antes</i>	'kaj	ˈa	ˈlus	ˈdus	ˈas	ˈtrus	ˈpur
<i>depois</i>	'kaj	ˈa	ˈlus	ˈdus	ˈas	ˈtrus	ˈpor
	1	2	3	4	5	6	7
<i>final</i>	'kaj	ˈa	ˈluz	ˈdu	ˈzas	ˈtrus	ˈpor
<i>final</i>	Cai	a	ˈluz	ˈdo	ˈs	ˈtros	ˈpor

**Regra 908**      acento      < 0,1‰

---

**ber:** /⟨ow⟩\$/ **msub:** “ow” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **mpts:** 1 **mton:** (tornar tônica)

MAT:33.2.4.3 porque era do ferro, ou

	1	2	3	4	5	6	7	8
<i>antes</i>	pur	ˈkɪ	ˈe	ˈra	ˈdu	ˈfe	ˈʁu	ˈow
<i>depois</i>	pur	ˈkɪ	ˈe	ˈra	ˈdu	ˈfe	ˈʁu	ˈow
	1	2	3	4	5	6	7	
<i>final</i>	pur	ˈkʲe	ˈra	ˈdu	ˈfe	ˈʁu	ˈow	
<i>final</i>	por	ˈque	e	ˈra	ˈdo	ˈfe	ˈrro	ˈou

## Camada 2

**Regra 1401** aférese 1,4‰ ★

**ber:** /([{aãeēĩĩoõõũũjw})(e)([sz][bdfptvmngk])/ **msub:** “\1\3” **bqsil:** 3

**bpar:** 2 **bton:** (indiferente, átona, indiferente)

**blim:** (final, inicial e não-final, indiferente) **mpts:** 4

**mreaj:** (nada, diminuir, nada) **mdel:** (nada, excluir, nada)

PIN.89.9.1 Eu estava como quem reprime atento

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'ew	es	'ta	va	'ko	mu	'kē	ʒe	'pri	mi	a	'tē	tu
<i>depois</i>	'ews	'ta	va	'ko	mu	'kē	ʒe	'pri	mi	a	'tē	tu	
<i>final</i>	'ews	'ta	va	'ko	mu	'kē	ʒe	'pri	mja	'tē	tu		
<i>final</i>	Eu	es	-ta-	va	-co-	mo	-quem	-re-	pri-	me	a-	ten-	to

**Regra 1404** síncope 0,2‰

**ber:** /([{kmpbt}][uʊ](lʊs?))/ **msub:** “\2\3” **bqsil:** 2 **bpar:** 1

**bton:** (pós-tônica, pós-tônica) **blim:** (não-inicial e não-final, final)

**bfron:** à direita **mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:4.25.1.1 Há quantos séculos já, plácida, airosa,

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'a	'kwā	tus	'sē	ku	'lus	'za	'pla	si	'da	'aj	'rɔ	'za
<i>depois</i>	'a	'kwā	tus	'sē	klus	'za	'pla	si	'da	'aj	'rɔ	'za	
<i>final</i>	'a	'kwā	tus	'sē	kluz	'za	'pla	si	'daj	'rɔ	'za		
<i>final</i>	Há	quan-	-tos	-sé-	culos	já,	plá-	-ci-	-da,	ai-	-ro-	-sa,	

**Regra 1407** síncope 0,2‰

**ber:** /([{bp})er(ā))/ **msub:** “\2r\3” **bqsil:** 2 **bpar:** 1

**bton:** (pré-tônica, tônica)

**blim:** (não-inicial e não-final, não-inicial e não-final) **bfron:** à esquerda

**mpts:** 10 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:2.38.4.4 Abriam lírios d'esperança.

	1	2	3	4	5	6	7	8	9
<i>antes</i>	a	'bri	āw	'li	rjus	des	pe	'rā	sa
<i>depois</i>	a	'bri	āw	'li	rjus	des	prā	sa	
<i>final</i>	a	'bri	āw	'li	rjuz	des	'prā	sa	
<i>final</i>	A-	bri-	am	-lí-	rios	-d'es-	peran-	-ça.	

## Camada 2 (cont.)

**Regra 1418**      sístole      0,2‰

---

**ber:** /([aæeēĩĩoõouū])ow)/([ɤʃsɒdfʒptvsʌŋlsrzmngk]+)i([ʊa]s?)/

**msub:** “\1\2\3” **bqsil:** 3 **bpar:** 1 **bton:** (átona, tônica, átona)

**blim:** (não-final, não-final, final) **bfron:** à direita **mpts:** 10

**mton:** (tornar tônica, tornar átona, nada)

DUR.4.37.5 Será de nigromancia um torpe ensaio,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	se	ˈra	ˈdɪ	ˈni	ˈgro	ˈmã	ˈsi	ˈa	ˈũ	ˈtɔr	ˈpɪ	ˈẽ	ˈsaj	u
<i>depois</i>	se	ˈra	ˈdɪ	ˈni	ˈgro	ˈmã	ˈsi	ˈa	ˈũ	ˈtɔr	ˈpɪ	ˈẽ	ˈsaj	u
	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	se	ˈra	ˈdɪ	ˈni	ˈgro	ˈmã	sɐũ	ˈtɔr	pɐẽ	ˈsaj	u			
<i>final</i>	Se	ˈrá	ˈde	ˈni	ˈgro	ˈman	ˈcia	um	ˈtor	ˈpe	en	ˈsai	ˈo	,

**Regra 1406**      síncope      0,1‰

---

**ber:** /(((t))a([ɤʃsɒdfʒptvsʌŋlsrzmngk]as?)))/ **msub:** “\2\3” **bqsil:** 2 **bpar:** 1

**bton:** (pós-tônica, pós-tônica) **blim:** (não-inicial e não-final, final)

**bfron:** à direita **mpts:** 3 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL.5.242.2.3 Juntei as pétalas delicadas, raras,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	zũ	ˈtej	ˈas	ˈpɛ	ˈta	ˈlas	ˈde	ˈli	ˈka	ˈdas	ˈka	ˈras
<i>depois</i>	zũ	ˈtej	ˈas	ˈpɛ	ˈtlas	ˈde	ˈli	ˈka	ˈdas	ˈka	ˈras	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	zũ	ˈtej	ˈas	ˈpɛ	ˈtlaz	ˈde	ˈli	ˈka	ˈdaz	ˈka	ˈras	
<i>final</i>	Jun	ˈtei	ˈas	ˈpé	ˈtalas	ˈde	ˈli	ˈca	ˈdas	ˈra	ˈras	,

**Regra 1420**      síncope      < 0,1‰

---

**ber:** /((kor([ɔo])))/ **msub:** “kr\2” **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, tônica)

**blim:** (inicial e não-final, não-inicial e não-final) **bfron:** à esquerda

**mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL.1.62.5.1 Tu és a coroa de virgínea palma,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	ˈtu	ˈɛs	ˈa	ˈko	ˈrɔ	ˈa	ˈdɪ	ˈvɪr	ˈzi	ˈnja	ˈpal	ˈma
<i>depois</i>	ˈtu	ˈɛs	ˈa	ˈkrɔ	ˈa	ˈdɪ	ˈvɪr	ˈzi	ˈnja	ˈpal	ˈma	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	ˈtu	ˈɛ	ˈza	ˈkrɔ	ˈa	ˈdɪ	ˈvɪr	ˈzi	ˈnja	ˈpal	ˈma	
<i>final</i>	Tu	ˈé	ˈs	ˈa	ˈcoro	ˈa	ˈde	ˈvɪr	ˈgí	ˈnea	ˈpal	ˈma

## Camada 2 (cont.)

**Regra 1415** síncope < 0,1%

**ber:** /(peri)/ **msub:** “pri” **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, pré-tônica)  
**blim:** (não-inicial e não-final, não-inicial e não-final) **bfron:** ambos  
**mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

CAM.1.29.3 Tantos climas e céus experimentados,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	tā	tus	kli	mas	i	sews	es	pe	ri	mē	ta	dus
<i>depois</i>	tā	tus	kli	mas	i	sews	es	pri	mē	ta	dus	
<i>final</i>	tā	tus	kli	ma	zi	sew	zes	pri	mē	ta	dus	
<i>final</i>	Tan	tos	cli	ma	se	céu	s	ex	peri	men	ta	dos,

**Regra 1403** paragoge < 0,1%

**ber:** /(b)/ **msub:** “bɪ” **bqsil:** 1 **bpar:** 1 **bton:** (átona)  
**blim:** (inicial e final) **bfron:** à direita **mpts:** 5 **mreaj:** (aumentar)

FIG:4.51.1.3 Sob noites tristíssimas, pressagas,

	1	2	3	4	5	6	7	8	9	10	11
<i>antes</i>	sub	no	tis	tris	ti	si	mas	pre	sa	gas	
<i>depois</i>	su	bɪ	no	tis	tris	ti	si	mas	pre	sa	gas
<i>final</i>	su	bɪ	no	tis	tris	ti	si	mas	pre	sa	gas
<i>final</i>	So	b	noi	tes	tris	tí	ssi	mas,	pre	ssa	gas,

**Regra 1409** síncope < 0,1%

**ber:** /(feli)/ **msub:** “fli” **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, pré-tônica)  
**blim:** (inicial e não-final, não-inicial e não-final) **bfron:** à esquerda  
**mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:1.94.4.2 dia, noite de dor ou felicidade,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	di	a	no	ti	dɪ	dɔ	row	feli	si	da	dɪ	
<i>depois</i>	di	a	no	ti	dɪ	dɔ	row	fli	si	da	dɪ	
<i>final</i>	di	a	no	ti	dɪ	dɔ	row	fli	si	da	dɪ	
<i>final</i>	di	a,	noi	te	de	do	r	ou	feli	ci	da	de,

## Camada 2 (cont.)

**Regra 1402** aférese < 0,1‰

**ber:** /([e])([sz][bdfptvmngk])/ **msub:** “\2” **bqsil:** 2 **bpar:** 1

**bton:** (átona, indiferente)

**blim:** (inicial e não-final, não-inicial e não-final) **bfron:** à esquerda

**mpts:** 15 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:5.189.3.4 “Estuda a lição!”

	1	2	3	4	5	6
<i>antes</i>	es	'tu	da	a	li	'sāw
<i>depois</i>	'stu	da	a	li	'sāw	
	1	2	3	4		
<i>final</i>	'stu	da	li	'sāw		
<i>final</i>	“Estu- da a li- ção!”					

**Regra 1408** síncope < 0,1‰

**ber:** /([fer]([eē]))/ **msub:** “fr\2” **bqsil:** 2 **bpar:** 1

**bton:** (pré-tônica, indiferente)

**blim:** (não-inicial e não-final, não-inicial e não-final) **bfron:** à esquerda

**mpts:** 8 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:1.75.19.6 levando oferenda seleta.

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	le	'vā	du	o	fe	'rē	da	se	'lɛ	ta
<i>depois</i>	le	'vā	du	o	'frē	da	se	'lɛ	ta	
	1	2	3	4	5	6	7	8		
<i>final</i>	le	'vā	dwo	'frē	da	se	'lɛ	ta		
<i>final</i>	le- van- do o- feren- da se- le- ta.									

**Regra 1426** síncope < 0,1‰

**ber:** /([p[eɛ])([aʊ]s?)/ **msub:** “p\2” **bqsil:** 2 **bpar:** 1

**bton:** (indiferente, átona) **blim:** (inicial e não-final, não-inicial e final)

**bfron:** ambos **mpts:** 10 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

**mton:** (tornar átona, nada)

SIL:5.55.1.1 Pela Natureza ressoa

	1	2	3	4	5	6	7	8	9	
<i>antes</i>	'pɛ	la	'na	tu	'rɛ	za	'ɛ	sɔ	a	
<i>depois</i>	pla	'na	tu	'rɛ	za	'ɛ	sɔ	a		
	1	2	3	4	5	6	7	8		
<i>final</i>	pla	'na	tu	'rɛ	za	'ɛ	sɔ	a		
<i>final</i>	Pela-Na-tu-re-za-re-sso-a									

## Camada 2 (cont.)

**Regra 1428** epêntese < 0,1‰

**ber:** /(kt|gn)/ **msub:** “XrY” **bqsil:** 2 **bpar:** 1 **blim:** (não-final, não-inicial)  
**mpts:** 10 **mreaj:** (aumentar, nada) **mmap:** (em 1 inserir 0 com mapa [kg] → [kg], em 3 inserir 1 com mapa [tn] → [tn])

SIL:5.184.4.3 Saúda o forte invicto Brasil.

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	sa	'u	da	'u	'fɔr	'tɪ	'vɪ	tu	'bra	'zil		
<i>depois</i>	sa	'u	da	'u	'fɔr	'tɪ	'vi	kɪ	tu	'bra	'zil	
	1	2	3	4	5	6	7	8	9	10		
<i>final</i>	sa	'u	du	'fɔr	tɪ	'vi	kɪ	tu	'bra	'zil		
<i>final</i>	Sa	-ú	-da	o	-for	-te	in	-vi	-ct-	o	Bra	-sil.

**Regra 1414** síncope < 0,1‰

**ber:** /ʁe(ver)/ **msub:** “ʁevr” **bqsil:** 3 **bpar:** 1  
**bton:** (pré-tônica, pré-tônica, tônica)  
**blim:** (inicial e não-final, não-inicial e não-final, não-inicial e não-final)  
**mpts:** 6 **mreaj:** (nada, diminuir, aumentar) **mdel:** (nada, excluir, nada)

MAT:38.3.3.1 Sem débita reverência seis andores,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	'sɛ	'dɛ	'bi	'ta	'ʁe	ve	'rɛ	'sja	'sejs	ã	'dɔ	'ris
<i>depois</i>	'sɛ	'dɛ	'bi	'ta	'ʁe	'vrɛ	'sja	'sejs	ã	'dɔ	'ris	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	'sɛ	'dɛ	'bi	'ta	'ʁe	'vrɛ	'sja	'sej	zã	'dɔ	'ris	
<i>final</i>	Sem	-dé	-bi	-ta	-re	-verên	-cia	-sei	-s	-an	-do	-res,

**Regra 1424** síncope < 0,1‰

**ber:** /(teri)/ **msub:** “tri” **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, pré-tônica)  
**blim:** (não-final, não-final) **bfron:** ambos **mpts:** 6  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

FIG:4.68.1.1 Quem morre solta, da misteriosa argila,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	'kɛ	'mɔ	'ʁɪ	'sɔ	'ta	'da	'mis	'te	'ri	'ɔ	'za	'ar	'zi	'la
<i>depois</i>	'kɛ	'mɔ	'ʁɪ	'sɔ	'ta	'da	'mis	'tri	'ɔ	'za	'ar	'zi	'la	
	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	'kɛ	'mɔ	'ʁɪ	'sɔ	'ta	'da	'mis	'trɔ	'zar	'zi	'la			
<i>final</i>	Quem	-mo	-re	-sol	-ta,	da	-mis	-terio	-sa	-ar	-gi	-la,		

## Camada 2 (cont.)

**Regra 1425** síncope < 0,1‰

**ber:** /(para)/ **msub:** “pra” **bqsil:** 2 **bpar:** 1 **bton:** (indiferente, átona)  
**blim:** (inicial e não-final, não-inicial e final) **bfron:** ambos **mpts:** 15  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

FIG:2.9.7.1 Pediste para que eu voltasse

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	pe	dis	tri	pa	ra	ki	ew	vol	ta	si
<i>depois</i>	pe	dis	tri	pra	ki	ew	vol	ta	si	
	1	2	3	4	5	6	7	8		
<i>final</i>	pe	dis	tri	pra	kjew	vol	ta	si		
<i>final</i>	Pe	dis	te	para	que eu	vol	ta	sse		

**Regra 1405** síncope < 0,1‰

**ber:** /(((p)u)([ɸsɓdf3ptvsʌŋlsrzmngk]as?))/ **msub:** “2\3” **bqsil:** 2 **bpar:** 1  
**bton:** (pós-tônica, pós-tônica) **blim:** (não-inicial e não-final, final)  
**bfron:** à direita **mpts:** 3 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:5.191.12.4 Do céu à cúpula anilada,

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	du	sew	a	ku	pu	la	a	ni	la	da
<i>depois</i>	du	sew	a	ku	pla	a	ni	la	da	
	1	2	3	4	5	6	7	8		
<i>final</i>	du	sew	a	ku	pla	ni	la	da		
<i>final</i>	Do	céu	à	cú	pula	a	ni	la	da,	

**Regra 1410** síncope < 0,1‰

**ber:** /(ãdesẽ)/ **msub:** “ãds'ẽ” **bqsil:** 3 **bpar:** 1  
**bton:** (pré-tônica, pré-tônica, tônica) **blim:** (não-inicial e não-final, não-inicial e não-final, não-inicial e não-final) **bfron:** à direita **mpts:** 6  
**mreaj:** (nada, diminuir, nada) **mdel:** (nada, excluir, nada)

FIG:3.34.1.3 De azul celeste, e como as resplandecências

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>antes</i>	di	a	zul	se	les	ti	i	ko	mu	as	wes	plã	de	sẽ	sjas
<i>depois</i>	di	a	zul	se	les	ti	i	ko	mu	as	wes	plã	dsẽ	sjas	
	1	2	3	4	5	6	7	8	9	10	11				
<i>final</i>	dja	zul	se	les	ti	ko	mwaz	wes	plã	dsẽ	sjas				
<i>final</i>	De	a	zul	ce	les	te,	e	co	mo	as	res	plan	decên	cias	



## Camada 2 (cont.)

**Regra 1413** síncope < 0,1‰

**ber:** /(pere)/ **msub:** “pre” **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, pré-tônica)  
**blim:** (inicial e não-final, não-inicial e não-final) **bfron:** à esquerda  
**mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:5.133.3.14 o rosto peregrino

	1	2	3	4	5	6	7
<i>antes</i>	u	ʔ	ɔs	tu	pe	re	gri nu
<i>depois</i>	u	ʔ	ɔs	tu	pre	gri	nu

	1	2	3	4	5	6
<i>final</i>	u	ʔ	ɔs	tu	pre	gri nu
<i>final</i>	o	ros	to	pere	gri	no

**Regra 1416** síncope < 0,1‰

**ber:** /(piri)/ **msub:** “pri” **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, pré-tônica)  
**blim:** (inicial e não-final, não-inicial e não-final) **bfron:** à esquerda  
**mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:1.44.2.3 a noite acende pirilampos,

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	a	no	ti	a	sê	dri	pi	ri	lã	pus
<i>depois</i>	a	no	ti	a	sê	dri	pri	lã	pus	

	1	2	3	4	5	6	7	8
<i>final</i>	a	no	tja	sê	dri	pri	lã	pus
<i>final</i>	a	noi	te a	cen	de	piri	lam	pos,

**Regra 1417** síncope < 0,1‰

**ber:** /(espiri)/ **msub:** “spri” **bqsil:** 3 **bpar:** 1  
**bton:** (pré-tônica, pré-tônica, pré-tônica)  
**blim:** (inicial, não-inicial e não-final, não-inicial e não-final)  
**bfron:** ambos **mpts:** 6 **mreaj:** (nada, diminuir, diminuir)  
**mdel:** (nada, excluir, excluir)

FIG:4.15.4.2 A própria luz do sol, espiritualizada, encerra

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
<i>antes</i>	a	pro	prja	lus	du	sow	es	pi	ri	tu	a	li	za	da	ẽ	se	ka
<i>depois</i>	a	pro	prja	lus	du	sow	spri	tu	a	li	za	da	ẽ	se	ka		

	1	2	3	4	5	6	7	8	9	10	11	12	13	
<i>final</i>	a	pro	prja	lus	du	sow	spri	twa	li	za	dẽ	se	ka	
<i>final</i>	A	pró	pria	luz	do	sol,	espiri	tua	li	za	da,	en	ce	rra

## Camada 2 (cont.)

**Regra 1427** síncope < 0,1‰

**ber:** /(der)/ **msub:** “dr” **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, tônica)

**blim:** (não-final, não-final) **bfron:** à esquerda **mpts:** 5

**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:5.160.6.2 reis poderosos

	1	2	3	4	5
<i>antes</i>	'	ɛjs	po	de	'rɔ
					zus
<i>depois</i>	'	ɛjs	po	dro	zus
	1	2	3	4	
<i>final</i>	'	ɛjs	po	'dro	zus
<i>final</i>	reis	po	dero	sos	

**Regra 1411** síncope < 0,1‰

**ber:** /(kula)/ **msub:** “kl'a” **bqsil:** 2 **bpar:** 1 **bton:** (átona, tônica)

**blim:** (não-inicial e não-final, não-inicial e não-final) **bfron:** ambos

**mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

SIL:1.92.1.1 Por sob o véu d'ímaculada brancura

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	pur	sub	u	'vew	di	ma	ku	'la	da	brã	'ku	ra
<i>depois</i>	pur	sub	u	'vew	di	ma	kla	da	brã	'ku	ra	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	pur	sub	u	'vew	di	ma	kla	da	brã	'ku	ra	
<i>final</i>	Por	sob	o	véu	d'i	ma	cula	da	bran	cu	ra	

**Regra 1412** síncope < 0,1‰

**ber:** /(kizi)/ **msub:** “kzi” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona)

**blim:** (não-inicial e não-final, não-inicial e não-final) **bfron:** ambos

**mpts:** 6 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

FIG:4.29.3.3 É por demais a inquisicional tortura

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	'	ɛ	pur	de	'maj	s	a	ĩ	ki	zi	si	o	'nal	tor
														'tu
<i>depois</i>	'	ɛ	pur	de	'maj	s	a	ĩ	kzi	si	o	'nal	tor	'tu
														ra
	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	'	ɛ	pur	de	'maj	z	a	ĩ	kzi	sjo	'nal	tor	'tu	ra
<i>final</i>	É	por	de	mai	s	a	in	quisi	cio	nal	tor	'tu	ra	

## Camada 2 (cont.)

**Regra 1419**      diástole       $< 0,1\%$

ber: /(r)(ia)(m)/ msub: “\1i'a\3” bqsil: 3 bpar: 2

**bton:** (tônica, átona, átona) **blim:** (não-final, não-final, final)

**bfron:** indiferente   **mpts:** 10   **mton:** (tornar átona, tornar tônica, nada)

FIG:4.39.2.1 E teríamos gozado o que gozamos

antes 1 2 3 4 5 6 7 8 9 10 11 12 13  
 depois 1 te ri a mus go za du u ki go zã mus  
 final 1 2 3 r ja muz go za du ki go zã mus  
 final E te ria mos go za do o que go za mos

**Regra 1421** síncope < 0,1‰

**ber:** /(kero)/ **msub:** "kro" **bqsil:** 2 **bpar:** 1 **bton:** (pré-tônica, tônica)

**blim:** (não-inicial e não-final, não-inicial e não-final) **bfron:** à esquerda

**mpts:** 6   **mreaj:** (nada, disminuir)   **mdel:** (nada, excluir)

SIL:5.126.37.6 úlceras feias, asquerosas,

antes 1 2 3 4 5 6 7 8 9  
 depois 1 2 3 4 5 6 7 8 9  
 final 1 2 3 4 5 6 7 8  
 final 1 2 3 4 5 6 7 8

**Regra 1422**      apócope       $< 0,1\%$

**ber:** /(marm[OU]rɪ)/   **msub:** “m'armor”   **bqsil:** 3   **bpar:** 1

**bton:** (tônica, átona, átona)

**blim:** (inicial e não-final, não-inicial e não-final, não-inicial e final)

**bfron:** ambos   **mpts:** 10   **mreaj:** (nada, nada, diminuir)

**mdel:** (nada, nada, excluir)

MAT:38.21.1.46 deste mármore, deste roble,

antes des- | ti- | mar- | mu- | ri- | des- | ti- | ro- | ble-  
depois des- | ti- | mar- | mur- | des- | ti- | ro- | ble-  
final des- | ti- | mar- | mur- | des- | ti- | ro- | ble-  
final des- | te- | mar- | more- | des- | te- | ro- | ble-

## Camada 2 (cont.)

**Regra 1423**      apócope      < 0,1%

**ber:** /is(ɪ)/   **msub:** “is”   **bqsil:** 2   **bpar:** 1   **bton:** (pós-tônica, pós-tônica)

**blim:** (não-inicial e não-final, final)   **bfron:** à direita   **mpts:** 6

**mreaj:** (nada, diminuir)   **mdel:** (nada, excluir)

FIG:2.15.20.2 E o seu peito um cálice de licor divino.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>antes</i>	ɪ	u	ˈsɛw	ˈpej	ˈtu	ũ	ˈka	li	ˈsɪ	ˈdɪ	li	ˈkɔɾ	ˈdi	ˈvi	ˈnu
<i>depois</i>	ɪ	u	ˈsɛw	ˈpej	ˈtu	ũ	ˈka	li	ˈsɪ	ˈdɪ	li	ˈkɔɾ	ˈdi	ˈvi	ˈnu
<i>final</i>	j	u	ˈsɛw	ˈpej	ˈtwũ	ˈka	li	ˈzɪ	ˈdɪ	li	ˈkɔɾ	ˈdi	ˈvi	ˈnu	
<i>final</i>	E	o	ˈsɛw	ˈpej	ˈtu	ũ	ˈka	li	ˈsɪ	ˈdɪ	li	ˈkɔɾ	ˈdi	ˈvi	ˈnu

**Regra 1429**      epêntese      < 0,1%

**ber:** /(bs)/   **msub:** “.XɪY”   **bqsil:** 1   **bpar:** 1   **blim:** (não-final)   **mpts:** 6

**mreaj:** (aumentar)   **mmap:** (em 1 inserir 0 com mapa [b] → [b], em 3 inserir 1 com mapa [s] → [s])

SIL:5.61.1.7 jamais obscurecidas

	1	2	3	4	5	6	7	8
<i>antes</i>	ʒ	a	ˈmaj	ˈobs	ku	ˈre	ˈsi	ˈdas
<i>depois</i>	ʒ	a	ˈmaj	ˈo	ˈbis	ku	ˈre	ˈsi
<i>final</i>	ʒ	a	ˈmaj	ˈzo	ˈbis	ku	ˈre	ˈsi
<i>final</i>	j	a	ˈmai	ˈs	ˈo	ˈbs	ˈcu	ˈre

**Regra 1430**      epêntese      < 0,1%

**ber:** /(bm)/   **msub:** “.XɪY”   **bqsil:** 2   **bpar:** 1   **blim:** (não-final, não-final)

**mpts:** 6   **mreaj:** (aumentar, nada)   **mmap:** (em 1 inserir 0 com mapa [b] → [b], em 3 inserir 1 com mapa [m] → [m])

SIL:5.115.1.9 foi batel submergido

	1	2	3	4	5	6	7	8
<i>antes</i>	ˈfoj	ˈba	ˈtɛw	ˈsub	ˈmer	ˈzi	ˈdu	
<i>depois</i>	ˈfoj	ˈba	ˈtɛw	ˈsu	ˈbɪ	ˈmer	ˈzi	ˈdu
<i>final</i>	ˈfoj	ˈba	ˈtɛw	ˈsu	ˈbɪ	ˈmer	ˈzi	ˈdu
<i>final</i>	foi	ˈba	ˈtel	ˈsu	ˈbm	ˈer	ˈgi	ˈdo

## Camada 3

**Regra 1105** sinérese 13,1‰ ★

**ber:** /([ei])([aeouãẽõũœ])/ **msub:** "j\2" **bqsil:** 2 **bpar:** 1  
**bton:** (átona, tônica) **blim:** (não-final, indiferente) **mpts:** 1  
**mdel:** (nada, excluir)

MAG.43.10.1 Ao som das vagas do Oceano

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	aw	sõ	das	va	gas	du	o	se	ã	nu
<i>depois</i>	aw	sõ	das	va	gas	du	o	sjã	nu	

	1	2	3	4	5	6	7	8
<i>final</i>	aw	sõ	daz	va	gaz	dwo	sjã	nu
<i>final</i>	Ao	som	das	va	gas	do	O	cea-no

**Regra 1101** sinérese 10,6‰ ★

**ber:** /([ei])([aeiouãẽõũœ])/ **msub:** "j\2" **bqsil:** 2 **bpar:** 1  
**bton:** (pré-tônica, pré-tônica) **blim:** (não-final, não-final) **mpts:** 0  
**mdel:** (nada, excluir)

PIN.16.16.3 Entre eles, da alma na expansão piedosa;

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>antes</i>	ẽ	tri	e	lis	da	al	ma	na	es	pã	sãw	pi	e	d	za
<i>depois</i>	ẽ	tri	e	lis	da	al	ma	na	es	pã	sãw	pje	d	za	

	1	2	3	4	5	6	7	8	9	10	11				
<i>final</i>	ẽ	trje	liz	dal	ma	naes	pã	sãw	pje	d	za				
<i>final</i>	En	tre	e	les	da	al	ma	na	ex	pan	são	pie	do	sa	

**Regra 1106** sinérese 5,9‰

**ber:** /([ou])([aoiãẽĩãõœ])/ **msub:** "w\2" **bqsil:** 2 **bpar:** 1  
**bton:** (átona, tônica) **blim:** (não-final, indiferente) **mpts:** 0  
**mdel:** (nada, excluir)

ANJ.25.74.4 À ruína vegetal dos lírios secos.

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	a	ku	i	na	ve	ze	tal	dus	li	rjus	se	kus
<i>depois</i>	a	kw	i	na	ve	ze	tal	dus	li	rjus	se	kus

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	a	kw	i	na	ve	ze	tal	duz	li	rjus	se	kus
<i>final</i>	À	ruí	na	ve	ge	tal	dos	lí	rios	se	cos	

## Camada 3 (cont.)

**Regra 1102** sinérese 2,7‰ ★

**ber:** /([ou])([eēaeo])/) **msub:** “w2” **bqsil:** 2 **bpar:** 1

**bton:** (pré-tônica, pré-tônica) **blim:** (não-final, não-final) **mpts:** 0

**mdel:** (nada, excluir)

CAM.10.133.6 Lágrimas no licor coalhado e enxuto

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	'la	gri	mas	nu	li	'kɔr	ko	a	'ʎa	du	ɪ	ẽ	'ʃu	tu
<i>depois</i>	'la	gri	mas	nu	li	'kɔr	kwa	'ʎa	du	ɪ	ẽ	'ʃu	tu	
<i>final</i>	'la	gri	maz	nu	li	'kɔr	kwa	'ʎa	dwjẽ	'ʃu	tu			
<i>final</i>	Lá	gri	mas	no	li	cor	coa	lha	do	e	en	xu	to	

**Regra 1107** sinérese 1,6‰ ★

**ber:** /([a])([ĩĩõ])/) **msub:** “aX” **bqsil:** 2 **bpar:** 1 **bton:** (átona, tônica)

**blim:** (não-final, indiferente) **mpts:** 8 **mdel:** (nada, excluir)

**mmap:** (em 1 inserir 1 com mapa [ĩĩõ] → [ĩĩõ̃])

FIG.1.25.4.2 Aonde estarás sonhando? E aonde estará florindo

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
<i>antes</i>	'aõ	dɪ	es	ta	'ras	so	'nã	du	ɪ	a	'õ	dɪ	es	ta	'ra	flo	'rĩ	du
<i>depois</i>	'aõ	dɪ	es	ta	'ras	so	'nã	du	ɪ	'aõ	dɪ	es	ta	'ra	flo	'rĩ	du	
<i>final</i>	'aõ	djes	ta	'ras	so	'nã				'dwjaõ	djes	ta	'ra	flo	'rĩ	du		
<i>final</i>	Aon	-de	es	-ta	-rás	so	-nhan	-do?	E	aon	-de	es	-ta	-rá	flo	-rin	-do	

**Regra 1123** sinérese 1,2‰

**ber:** /([ua])/ **msub:** “wa” **bqsil:** 2 **bpar:** 1 **bton:** (indiferente, átona)

**blim:** (não-final, final) **mpts:** 4 **mdel:** (nada, excluir)

**mton:** (tornar átona, nada)

CAM.3.85.3 E que em sua vida já se exprimentara,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ɪ	kɪ	ẽ	'su	a	'vi	da	'za	sɪ	es	pri	mẽ	'ta	ra
<i>depois</i>	ɪ	kɪ	ẽ	swa	'vi	da	'za	sɪ	es	pri	mẽ	'ta	ra	
<i>final</i>	ɪ	kjẽ	swa	'vi	da	'za	sjes	pri	mẽ	'ta	ra			
<i>final</i>	E	que	em	sua	vi	da	já	se	ex	pri	men	-ta	-ra,	

### Camada 3 (cont.)

**Regra 1117** sinérese 0,8‰

**ber:** /([εe]ē)/ **msub:** “ē” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)  
**blim:** (não-final, indiferente) **mpts:** 0 **mreaj:** (nada, diminuir)  
**mdel:** (nada, excluir)

MAT:9.10.28.7 não se lhes dêem de ninguém,

	1	2	3	4	5	6	7	8
<i>antes</i>	'nāw	sɪ	ʎɪs	de	ē	dɪ	nĩ	gē
<i>depois</i>	'nāw	sɪ	ʎɪs	dē	dɪ	nĩ	gē	
	1	2	3	4	5	6	7	
<i>final</i>	'nāw	sɪ	ʎɪz	dē	dɪ	nĩ	gē	
<i>final</i>	não	se	lhes	dêem	de	nin	guém,	

**Regra 1103** sinérese 0,4‰

**ber:** /([a([eo]))/ **msub:** “aX” **bqsil:** 2 **bpar:** 1  
**bton:** (pré-tônica, pré-tônica) **blim:** (não-final, não-final) **mpts:** 1  
**mdel:** (nada, excluir) **mmap:** (em 1 inserir 1 com mapa [eo] → [jw])

MAT:12.15.9.9 como Faetonte rodara,

	1	2	3	4	5	6	7	8	9
<i>antes</i>	'ko	mu	fa	e	tõ	tɪ	ʁo	da	ra
<i>depois</i>	'ko	mu	faj	tõ	tɪ	ʁo	da	ra	
	1	2	3	4	5	6	7	8	
<i>final</i>	'ko	mu	faj	tõ	tɪ	ʁo	da	ra	
<i>final</i>	co	mo	Fae	ton	te	ro	da	ra,	

**Regra 1112** sinérese 0,4‰

**ber:** /([i][ʊ])/ **msub:** “XY” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)  
**blim:** (não-final, indiferente) **mpts:** 3 **mdel:** (nada, excluir)  
**mmap:** (em 0 inserir 0 com mapa [æi] → [æi], em 1 inserir 1 com mapa [aoujēō] → [aowjēō])

DUR.3.4.4 Clara luz de um principio sempiterno.

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'kla	ra	lus	dɪ	ũ	prĩ	sɪ	pi	u	sē	pi	ter	nu
<i>depois</i>	'kla	ra	lus	dɪ	ũ	prĩ	sɪ	piw	sē	pi	ter	nu	
	1	2	3	4	5	6	7	8	9	10	11		
<i>final</i>	'kla	ra	luz	djũ	prĩ	sɪ	piw	sē	pi	ter	nu		
<i>final</i>	Cl	a	ra	luz	de	um	prin	ci	pio	sem	pi	ter	no.

## Camada 3 (cont.)

**Regra 1109** sinérese 0,3‰

**ber:** /([æe])[ɔ]/ **msub:** "XY" **bqsil:** 2 **bpar:** 1 **btón:** (tônica, átona)

**blim:** (não-final, indiferente) **mpts:** -1 **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [æe] → [æe], em 1 inserir 1 com mapa [aouĩẽõ] → [ə̞wĩẽõ])

ANJ.1.1.4 Larva de caos telúrico, procedo

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	'lar	va	dĩ	'ka	us	te	'lu	ri	ku	pro	'se	du
<i>depois</i>	'lar	va	dĩ	'kaws	te	'lu	ri	ku	pro	'se	du	

	1	2	3	4	5	6	7	8	9	10	11
<i>final</i>	'lar	va	dĩ	'kaws	te	'lu	ri	ku	pro	'se	du
<i>final</i>	Lar	va	de	caos	te	lú	ri	co	pro	ce	do

**Regra 1128** diérese 0,3‰

**ber:** /([aeio])([w])/ **msub:** "\1.X" **bqsil:** 1 **bpar:** 2 **btón:** (pré-tônica)

**blim:** (não-final) **mpts:** 12 **mmap:** (em 1 inserir 0 com mapa [jw] → [iu])

GON.10.9.56 Os ache miudinhos, sempre servem.

	1	2	3	4	5	6	7	8	9	10	11
<i>antes</i>	us	'a	ʃĩ	miw	'di	jus	'sẽ	prĩ	'ser	vẽ	
<i>depois</i>	us	'a	ʃĩ	mi	u	'di	jus	'sẽ	prĩ	'ser	vẽ

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	u	'za	ʃĩ	mi	u	'di	jus	'sẽ	prĩ	'ser	vẽ	
<i>final</i>	O	s	a	che	mi	u	di	nhos	sem	pre	ser	vem.

**Regra 1127** diérese 0,2‰

**ber:** /([aeou])([j])/ **msub:** "\1.X" **bqsil:** 1 **bpar:** 2 **btón:** (pré-tônica)

**blim:** (não-final) **mpts:** 15 **mmap:** (em 1 inserir 0 com mapa [jw] → [iu])

CAM.4.99.1 — "Já que nesta gostosa vaidade

	1	2	3	4	5	6	7	8	9	10	11
<i>antes</i>	'za	kĩ	'nes	ta	gos	'to	za	vaj	'da	dĩ	
<i>depois</i>	'za	kĩ	'nes	ta	gos	'to	za	va	i	'da	dĩ

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	'za	kĩ	'nes	ta	gos	'to	za	va	i	'da	dĩ	
<i>final</i>	—	Já	que	nes	ta	gos	to	sa	va	i	da	de



### Camada 3 (cont.)

**Regra 1129** diérese 0,1‰

**ber:** /([auo])(j)/ **msub:** “\1.i” **bqsil:** 1 **bpar:** 2 **bton:** (tônica) **mpts:** 15

PIN.17.33.2 Em circuitos largos sem ter pressa:

	1	2	3	4	5	6	7	8	9	10	11									
<i>antes</i>	ẽ	sir	'	kuj	tus	'	lar	gus	'	sẽ	ter	'	pre	'	sa					
<i>depois</i>	ẽ	sir	'	ku	i	tus	'	lar	gus	'	sẽ	ter	'	pre	'	sa				
	1	2	3	4	5	6	7	8	9	10	11									
<i>final</i>	ẽ	sir	'	ku	i	tuz	'	lar	gus	'	sẽ	ter	'	pre	'	sa				
<i>final</i>	Em	cir	-	cu	-	i	-	tos	'	lar	-	gos	'	sem	'	ter	'	pre	-	ssa

**Regra 1119** sinérese 0,1‰

**ber:** /([iɪ])([ʊa])/ **msub:** “j2” **bqsil:** 2 **bpar:** 1

**bton:** (pós-tônica, pós-tônica) **blim:** (não-inicial, não-inicial) **mpts:** -1

**mdel:** (nada, excluir)

CAM.10.103.3 Que se injuria de usar fundido cobre

	1	2	3	4	5	6	7	8	9	10	11	12	13	14						
<i>antes</i>	kɪ	sɪ	'	tɪ	'	zu	ri	a	dɪ	u	'	zar	fũ	'	di	du	'	kɔ	'	brɪ
<i>depois</i>	kɪ	sɪ	'	tɪ	'	zu	rja	dɪ	u	'	zar	fũ	'	di	du	'	kɔ	'	brɪ	
	1	2	3	4	5	6	7	8	9	10	11									
<i>final</i>	kɪ	sɪ	'	zu	rja	dju	'	zar	fũ	'	di	du	'	kɔ	'	brɪ				
<i>final</i>	Que	se	in	-	ju	-	ria	de	u	-	sar	fun	-	di	-	do	co	-	bre	

**Regra 1114** sinérese 0,1‰

**ber:** /([aɔ])([ẽ])/ **msub:** “XY” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)

**blim:** (não-final, indiferente) **mpts:** 2 **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [aɔ] → [aɔ], em 1 inserir 1 com mapa [aouĩẽõ] → [aɔwĩẽõ])

MAT:36.2.27.2 já caem no pecado velho,

	1	2	3	4	5	6	7	8	9									
<i>antes</i>	'	ʒa	'	'ka	'	ẽ	'	nu	'	pe	'	'ka	'	du	'	'vɛ	'	ʎu
<i>depois</i>	'	ʒa	'	'ka	'	ẽ	'	nu	'	pe	'	'ka	'	du	'	'vɛ	'	ʎu
	1	2	3	4	5	6	7	8										
<i>final</i>	'	ʒa	'	'ka	'	ẽ	'	nu	'	pe	'	'ka	'	du	'	'vɛ	'	ʎu
<i>final</i>	já	caem	no	pe	-	ca	-	do	ve	-	lho	,						

## Camada 3 (cont.)

**Regra 1125** sinérese < 0,1%

**ber:** /(uma)/ **msub:** “wa” **bqsil:** 2 **bpar:** 1 **bton:** (indiferente, átona)  
**blim:** (inicial e não-final, não-inicial e final) **bfron:** ambos **mpts:** 15  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir) **mton:** (tornar átona, nada)

FIG:1.29.1.3 De uma garoa marinha, uma garça marinha,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>antes</i>	dɪ	u	ma	ga	ro	a	ma	ri	na	u	ma	gar	sa	ma	ri	na
<i>depois</i>	dɪ	wa	ga	ro	a	ma	ri	na	u	ma	gar	sa	ma	ri	na	

	1	2	3	4	5	6	7	8	9	10	11	12	13			
<i>final</i>	djwa	ga	ro	a	ma	ri	nu	ma	gar	sa	ma	ri	na			
<i>final</i>	De	uma	ga	ro	a	ma	ri	nha	u	ma	gar	ça	ma	ri	nha	

**Regra 1118** sinérese < 0,1%

**ber:** /(i([ēāō]))/ **msub:** “j2” **bqsil:** 2 **bpar:** 1  
**bton:** (pós-tônica, pós-tônica) **blim:** (não-inicial, não-inicial) **mpts:** -2  
**mdel:** (nada, excluir)

ANJ.12.15.2 Ou, pelo menos, o ignis sapiens do Orco

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ow	pɛ	lu	me	nus	u	ig	nis	sa	pi	ēs	du	ɔr	ku
<i>depois</i>	ow	pɛ	lu	me	nus	u	ig	nis	sa	pjēs	du	ɔr	ku	

	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	ow	pɛ	lu	me	nu	z	wig	nis	sa	pjēz	dwɔr	ku		
<i>final</i>	Ou	pe	lo	me	no	s	o	ig	nis	sá	piens	do	Or	co

**Regra 1111** sinérese < 0,1%

**ber:** /(ɛa)/ **msub:** “ɛā” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, pós-tônica)  
**blim:** (não-inicial e não-final, final) **mpts:** 6 **mdel:** (nada, excluir)

PIN.14.36.3 Eneas do busto as partes bem dispostas.

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	e	ne	as	du	bus	tu	as	par	tis	bē	dis	pos	tas
<i>depois</i>	e	neqs	du	bus	tu	as	par	tis	bē	dis	pos	tas	

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	e	neqz	du	bus	twas	par	tiz	bē	dis	pos	tas	
<i>final</i>	E	neas	do	bus	to	as	par	tes	bem	dis	pos	tas

### Camada 3 (cont.)

**Regra 1108** sinérese < 0,1‰

**ber:** /(e[í])/ **msub:** “eX” **bqsil:** 2 **bpar:** 1 **bton:** (átona, tônica)

**blim:** (não-final, indiferente) **mpts:** 10 **mdel:** (nada, excluir)

**mmap:** (em 1 inserir 1 com mapa [i] → [j])

ANJ.1.30.1 Era a elégia panteísta do Universo,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
<i>antes</i>	ε	ra	a	e	l	ε	z	ja	p	ā	te	is	ta	du	u	ni	ver	su
<i>depois</i>	ε	ra	a	e	l	ε	z	ja	p	ā	tejs	ta	du	u	ni	ver	su	
	1	2	3	4	5	6	7	8	9	10	11							
<i>final</i>	ε	ra	ε	l	ε	z	ja	p	ā	tejs	ta	du	ni	ver	su			
<i>final</i>	E	ra	a	e	lé	gia	pan	teis	ta	do	U	ni	ver	so				

**Regra 1113** sinérese < 0,1‰

**ber:** /([i][o])/ **msub:** “XY” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)

**blim:** (não-final, indiferente) **mpts:** 5 **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [aei] → [aei], em 1 inserir 1 com mapa [aouĩẽõ] → [aowjẽõ])

SIL.1.128.1.21 veloz alcíone os ares atravessa.

	1	2	3	4	5	6	7	8	9	10	11	12	13			
<i>antes</i>	ve	l	os	al	si	o	ni	us	a	ris	a	tra	ve	sa		
<i>depois</i>	ve	l	os	al	si	o	ni	us	a	ris	a	tra	ve	sa		
	1	2	3	4	5	6	7	8	9	10	11					
<i>final</i>	ve	l	o	zal	si	o	nju	za	ri	za	tra	ve	sa			
<i>final</i>	ve	l	o	z	al	cío	ne	o	s	a	re	s	a	tra	ve	ssa

**Regra 1115** sinérese < 0,1‰

**ber:** /([iu][õẽ])/ **msub:** “XY” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)

**blim:** (não-final, indiferente) **mpts:** 6 **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [iu] → [iu], em 1 inserir 1 com mapa [aouĩẽõ] → [aowjẽõ])

ANJ.67.17.3 Afluem milhões de sóis, rubros, radiando...

	1	2	3	4	5	6	7	8	9	10	11	12				
<i>antes</i>	a	flu	ẽ	mi	l	õjs	d	i	s	ajs	ku	brus	ka	dj	ā	du
<i>depois</i>	a	flu	ẽ	mi	l	õjs	d	i	s	ajs	ku	brus	ka	dj	ā	du
	1	2	3	4	5	6	7	8	9	10	11					
<i>final</i>	a	flu	ẽ	mi	l	õjz	d	i	s	ajz	ku	bruz	ka	dj	ā	du
<i>final</i>	A	fluem	mi	lhões	de	sóis	ru	bro	s	ra	dian	do				

## Camada 3 (cont.)

**Regra 1120** sinérese < 0,1%

**ber:** /([ʊo](ɪ))/ **msub:** “w\2” **bqsil:** 2 **bpar:** 1

**bton:** (pós-tônica, pós-tônica) **blim:** (não-inicial, não-inicial) **mpts:** 0

**mdel:** (nada, excluir)

CAM.9.2.2 Arsínoe foi do Egípcio Ptolomeu,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	ar	si	nu	ɪ	foj	du	e	zip	sjɯ	pto	lo	'mew
<i>depois</i>	ar	si	nwɪ	foj	du	e	zip	sjɯ	pto	lo	'mew	
	1	2	3	4	5	6	7	8	9	10		
<i>final</i>	ar	si	nwɪ	foj	dwe	zip	sjɯ	pto	lo	'mew		
<i>final</i>	Ar	sí	noe	foi	do	E	gíp	cio	Pto	lo	meu,	

**Regra 1110** sinérese < 0,1%

**ber:** /([a]ɪ)/ **msub:** “XY” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)

**blim:** (não-final, indiferente) **mpts:** -1 **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [aæ] → [aæ], em 1 inserir 1 com mapa [aouĩẽõ] → [aṵwĩẽṵ])

MAT:36.4.1.40 gabe-a por esses poiaes.

	1	2	3	4	5	6	7	8	9
<i>antes</i>	'ga	bɪ	a	pur	'ɛ	sɪs	poj	'a	ɪs
<i>depois</i>	'ga	bɪ	a	pur	'ɛ	sɪs	poj	'a	js
	1	2	3	4	5	6	7		
<i>final</i>	'ga	bja	pu	'rɛ	sɪs	poj	'a	js	
<i>final</i>	ga	-be-a	-po-	r	e-	sses	-poi-	aes.	

**Regra 1116** sinérese < 0,1%

**ber:** /([ʊ]ɪ)/ **msub:** “uY” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)

**blim:** (não-final, indiferente) **mpts:** 3 **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [u] → [u], em 1 inserir 1 com mapa [aouĩẽõ] → [aṵwĩẽṵ])

ANJ.90.4.1 Restituo triunfalmente à esfera calma

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ɤs	ti	'tu	ɯ	trjũ	fal	'mẽ	tɪ	a	es	'fɛ	ra	'kal	ma
<i>depois</i>	ɤs	ti	'tuw	trjũ	fal	'mẽ	tɪ	a	es	'fɛ	ra	'kal	ma	
	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	ɤs	ti	'tuw	trjũ	fal	'mẽ	tjaɤs	'fɛ	ra	'kal	ma			
<i>final</i>	Res	-ti-	tuo	triun-	fal-	men-	te	à	es-	fɛ	ra	cal-	ma	

### Camada 3 (cont.)

**Regra 1121** sinérese < 0,1‰

**ber:** /(a)/ **msub:** “aj” **bqsil:** 2 **bpar:** 1 **bton:** (pós-tônica, pós-tônica)  
**blim:** (não-inicial, não-inicial) **mpts:** -1 **mdel:** (nada, excluir)

COS.117.132.1 De Europa, Dânae, Leda, e mais Almena,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14															
<i>antes</i>	d	i	e	w	r	o	p	a	d	ã	n	a	i	l	e	d	a	i	m	a	j	s	a	l	m	e	n	a	
<i>depois</i>	d	i	e	w	r	o	p	a	d	ã	n	a	j	i	l	e	d	a	i	m	a	j	s	a	l	m	e	n	a
	1	2	3	4	5	6	7	8	9	10	11																		
<i>final</i>	d	j	e	w	r	o	p	a	d	ã	n	a	j	i	l	e	d	i	m	a	j	z	a	l	m	e	n	a	
<i>final</i>	D	e	E	u	r	o	p	a	,D	â	n	a	e	,L	e	d	a	,e	m	a	i	s	A	l	m	e	n	a	,

**Regra 1126** sinérese < 0,1‰

**ber:** /(eɨ)/ **msub:** “ej” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)  
**blim:** (não-final, final) **bfron:** à direita **mpts:** 15  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

ANJ.25.41.1 Veio-lhe então como à fêmea vêm antojos,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14																														
<i>antes</i>	'	v	e	j	u	l	i	ẽ	'	t	ã	w	'	k	o	m	u	a	'	f	e	'	m	j	a	'	v	ẽ	ã	'	t	o	'	z	u	s								
<i>depois</i>	'	v	e	j	l	i	ẽ	'	t	ã	w	'	k	o	m	u	a	'	f	e	'	m	j	a	'	v	ẽ	ã	'	t	o	'	z	u	s									
	1	2	3	4	5	6	7	8	9	10	11																																	
<i>final</i>	'	v	e	j	l	i	ẽ	'	t	ã	w	'	k	o	m	w	a	'	f	e	'	m	j	a	'	v	ẽ	ã	'	t	o	'	z	u	s									
<i>final</i>	V	e	i	o	-	l	h	e	e	n	-	t	ã	o	-	c	o	-	m	o	-	à	-	f	ê	-	m	e	a	-	v	ê	m	-	a	n	-	t	o	-	j	o	s	,

**Regra 1130** diérese < 0,1‰

**ber:** /(ja)/ **msub:** “i.a” **bqsil:** 1 **bpar:** 1 **bton:** (pós-tônica) **mpts:** 15

SIL:5.40.1.8 São orquídeas formosas,

	1	2	3	4	5	6	7	8																			
<i>antes</i>	'	s	ã	w	'	o	r	'	k	i	'	d	j	a	s	'	f	o	r	'	m	o	'	z	a	s	
<i>depois</i>	'	s	ã	w	'	o	r	'	k	i	'	d	i	'	a	s	'	f	o	r	'	m	o	'	z	a	s
	1	2	3	4	5	6	7	8																			
<i>final</i>	'	s	ã	w	'	o	r	'	k	i	'	d	i	'	a	s	'	f	o	r	'	m	o	'	z	a	s
<i>final</i>	S	ã	o	-	r	q	u	i	-	d	e	a	-	s	-	f	o	-	r	-	m	o	-	s	a	s	,

## Camada 4

**Regra 2001**      redução      0,3‰

---

**ber:** /(<aw)/ **msub:** “a” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 7 **mreaj:** (diminuir)

CAM.2.86.2 Responde ao embaixador, que tanto estima:

antes 1 2 3 4 5 6 7 8 9 10 11 12 13 14  
 antes ʁes|põ|di|aw|ẽ|ba|ʃa|dõr|kɪ|tã|tu|es|ti|ma  
 depois ʁes|põ|di|a|ẽ|ba|ʃa|dõr|kɪ|tã|tu|es|ti|ma

final 1 2 3 4 5 6 7 8 9 10 11  
 final ʁes|põ|djaẽ|ba|ʃa|dõr|kɪ|tã|twes|ti|ma  
 final Res|pon|de|ao|em|ba|xa|dor|que|tan|to|es|ti|ma:

**Regra 2002**      redução      0,2‰

---

**ber:** /(<ow)/ **msub:** “u” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 6 **mreaj:** (diminuir)

FIG:1.188.13.2 Mocidade morreu, ou, então, se transformou.

antes 1 2 3 4 5 6 7 8 9 10 11 12 13  
 antes mo|si|da|di|mo|ʁew|ow|ẽ|tãw|si|trãs|for|mow  
 depois mo|si|da|di|mo|ʁew|u|ẽ|tãw|si|trãs|for|mow

final 1 2 3 4 5 6 7 8 9 10 11 12  
 final mo|si|da|di|mo|ʁew|wẽ|tãw|si|trãs|for|mow  
 final Mo|ci|da|de|mo|rreu|ou,en|tã|se|trans|for|mou.

**Regra 2003**      redução      < 0,1‰

---

**ber:** /(<tu)/ **msub:** “tʊ” **bqsil:** 1 **bpar:** 1 **bton:** (tônica)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 10 **mton:** (tornar átona)

MAT:10.28.3.2 Tu a vida deixas, eu a morte imploro

antes 1 2 3 4 5 6 7 8 9 10 11 12 13  
 antes tu|a|vi|da|dej|ʃas|ew|a|mõr|tɪ|plõr|u  
 depois tu|a|vi|da|dej|ʃas|ew|a|mõr|tɪ|plõr|u

final 1 2 3 4 5 6 7 8 9 10 11  
 final twa|vi|da|dej|ʃa|zew|a|mõr|tɪ|plõr|u  
 final Tu|a|vi|da|dei|xa|s,eu|a|mõr|te|im|plo|ro

## Camada 5

Regra 601	ectlipse	6,4‰	★
-----------	----------	------	---

ber: /(kõ)/ msub: “kʊ” bqsil: 1 bpar: 1 bton: (átona)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 1 **mton:** (tornar átona)

CAM.4.17.4 Senão co'os vossos fortes pais, e avôs?

1 2 3 4 5 6 7 8 9 10 11 12  
*antes* se | nãw | kō | us | vɔ | sus | for | tis | pajɔ | I | a | vos  
*depois* se | nãw | ku | us | vɔ | sus | for | tis | pajɔ | I | a | vos

1 2 3 4 5 6 7 8 9 10  
final se nãw kuz vo sus for tis paj zja vos  
final Se- não com os vo-ssos for-tes pai-s, e a- vôs?

**Regra 605**      eclipse      1,0‰

ber: /(ãw)/ msub: "a" bqsil: 1 bpar: 1 bton: (átona)

**blim:** (não-inicial e final) **bfron:** à direita **mpts:** 3 **mreaj:** (diminuir)

SIL:5.56.1.5 Voltam os ledos passarinhos,

1 2 3 4 5 6 7 8 9  
*antes* 'vɔl tãw us 'lɛ dus pa sa 'ri nus  
*depois* 'vɔl ta us 'lɛ dus pa sa 'ri nus

1 2 3 4 5 6 7 8  
*final* 'vol: tuz: 'le: dus: pa: sa: 'ri: nus  
*final* Vol: ta os: le: dos: pa: ssa: ri: nhos,

Regra 606	ectlipse	0,6‰
-----------	----------	------

ber: /([ẽõ])/ msub: "X" bqsil: 1 bpar: 1 bton: (átona)

**blim:** (não-inicial e final)   **bfron:** à direita   **mpts:** 4

**mmap:** (em 0 inserir 0 com mapa  $[\tilde{e}\tilde{o}] \rightarrow [IV]$ )

**CAM.7.28.7** Sobem à capitania; e toda a gente

antes 'sɔ bẽ a ka pi ta' ni a i 'tɔ da a 'zẽ ti  
depois 'sɔ bi a ka pi ta' ni a i 'tɔ da a 'zẽ ti

1 2 3 4 5 6 7 8 9 10 11  
*final* 'so bja ka pi ta ni i 'to da 'zẽ ti  
*final* So- bem à ca- pi- ta- nia- e to- da a gen- te

## Camada 5 (cont.)

**Regra 608**      ecllipse      < 0,1‰

---

**ber:** /(ua(s?))/ **msub:** “wa\2” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)

**blim:** (inicial, final) **bfron:** ambos **mpts:** -1 **mdel:** (nada, excluir)

**mton:** (tornar átona, nada)

MAT:25.10.1.8 de u'as agulhas ferrugentas,

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	dɪ	ˈu	as	a	ˈgu	ˈlas	fe	ku	ˈzẽ	tas
<i>depois</i>	dɪ	was	a	ˈgu	ˈlas	fe	ku	ˈzẽ	tas	
	1	2	3	4	5	6	7	8		
<i>final</i>	djwa	za	ˈgu	ˈlas	fe	ku	ˈzẽ	tas		
<i>final</i>	de u'as	a-	gu-	lhas	fe-	rru-	gen-	tas,		

**Regra 603**      ecllipse      < 0,1‰

---

**ber:** /(([nsk])ẽ)/ **msub:** “\2ɪ” **bqsil:** 1 **bpar:** 1 **bton:** (tônica)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 15 **mton:** (tornar átona)

FIG:2.12.1.3 Sem o sol do verão,

	1	2	3	4	5	6
<i>antes</i>	ˈsẽ	u	ˈsɔw	du	ve	ˈrãw
<i>depois</i>	sɪ	u	ˈsɔw	du	ve	ˈrãw
	1	2	3	4	5	
<i>final</i>	sju	ˈsɔw	du	ve	ˈrãw	
<i>final</i>	Sem o	sol	do	ve-	rão,	

**Regra 607**      ecllipse      < 0,1‰

---

**ber:** /(nãw)/ **msub:** “nɔ” **bqsil:** 1 **bpar:** 1 **bton:** (tônica)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 15 **mton:** (tornar átona)

FIG:2.24.2.2 E não há nada como as preces

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	ɪ	ˈnãw	ˈa	ˈna	ˈda	ˈko	ˈmu	as	ˈprɛ	sɪs
<i>depois</i>	ɪ	nu	ˈa	ˈna	ˈda	ˈko	ˈmu	as	ˈprɛ	sɪs
	1	2	3	4	5	6	7	8		
<i>final</i>	ɪ	ˈnwa	ˈna	ˈda	ˈko	ˈmwas	ˈprɛ	sɪs		
<i>final</i>	E	nã	hã	na-	da	co-	mo	as	pre-	ces



Camada 5 (*cont.*)

**Regra 609**      ecllipse      < 0,1‰

**ber:** /(ũ)/ **msub:** “ʊ” **bqsil:** 1 **bpar:** 1 **bton:** (átona)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 15

SIL:5.246.3.3 não teve um ecoo seu — ai —,

	1	2	3	4	5	6	7	8	9
<i>antes</i>	'nāw	'tɛ	vɪ	ũ	e	kɔ	u	'sew	'aj
<i>depois</i>	'nāw	'tɛ	vɪ	u	e	kɔ	u	'sew	'aj
	1	2	3	4	5	6	7		
<i>final</i>	'nāw	'tɛ	vjwe	'kɔ	u	'sew	'aj		
<i>final</i>	não	te	ve um e	co	o	seu	—	ai	

**Regra 602**      ecllipse      < 0,1‰

**ber:** /(tāw)/ **msub:** “taw” **bqsil:** 1 **bpar:** 1 **bton:** (tônica)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 6 **mton:** (tornar átona)

MAT:38.5.1.30 porque não sou tão infeliz,

	1	2	3	4	5	6	7	8
<i>antes</i>	pur	kɪ	'nāw	'sow	'tāw	'tɛ	'lis	
<i>depois</i>	pur	kɪ	'nāw	'sow	taw	'tɛ	'lis	
	1	2	3	4	5	6	7	
<i>final</i>	pur	kɪ	'nāw	'sow	twĩ	'fe	'lis	
<i>final</i>	por	que	não	sou	tão in	fe	liz,	

**Regra 610**      ecllipse      < 0,1‰

**ber:** /(nũ)/ **msub:** “nũ” **bqsil:** 1 **bpar:** 1 **bton:** (indiferente)

**blim:** (inicial e final) **bfron:** ambos **mpts:** 10 **mton:** (tornar átona)

SIL:5.196.2.2 Repousava num altar de flores e de luz;

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	ɛ	pow	'za	va	'nũ	'al	'tar	'dɪ	'flo	'rɪs	'ɪ	'dɪ	'lus
<i>depois</i>	ɛ	pow	'za	va	nu	'al	'tar	'dɪ	'flo	'rɪs	'ɪ	'dɪ	'lus
	1	2	3	4	5	6	7	8	9	10	11	12	
<i>final</i>	ɛ	pow	'za	va	nwal	'tar	'dɪ	'flo	'rɪ	'zɪ	'dɪ	'lus	
<i>final</i>	Re	pou	sa	va	num al	tar	de	flo	re	s e	de	luz;	

## Camada 5 (cont.)

**Regra 604**      eclipse      < 0,1‰

**ber:** /(ē)/   **msub:** “i”   **bqsil:** 1   **bpar:** 1   **bton:** (átona)   **blim:** (inicial e final)  
**bfron:** ambos   **mpts:** 15   **mton:** (tornar átona)

SIL:5.126.68.2 muito velava em oração:

	1	2	3	4	5	6	7	8	9
<i>antes</i>	'muj	tu	ve	'la	va	ẽ	o	ra	'sãw
<i>depois</i>	'muj	tu	ve	'la	va	i	o	ra	'sãw
	1	2	3	4	5	6	7		
<i>final</i>	'muj	tu	ve	'la	vjo	ra	'sãw		
<i>final</i>	mui-	to	ve-	la-	va	em	o-	ra-	ção:

## Camada 6

**Regra 1002**      elisão      216,3‰      ★

**ber:** /(a)/([aeiouɾɛɔãẽĩõũ])/   **msub:** “2”   **bqsil:** 2   **bpar:** 1  
**bton:** (átona, átona)   **blim:** (não-inicial e final, inicial)   **mpts:** -2  
**mreaj:** (nada, diminuir)   **mdel:** (nada, excluir)

PIN.23.31.3 Dizer quem sejas, não recusa esquivo”.

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	di	'zer	'kẽ	'se	zas	'nãw	'be	'ku	za	es	'ki	'vu
<i>depois</i>	di	'zer	'kẽ	'se	zas	'nãw	'be	'ku	zes	'ki	'vu	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	di	'zer	'kẽ	'se	zas	'nãw	'be	'ku	zes	'ki	'vu	
<i>final</i>	Di-	zer	quem	se-	jas,	não	re-	cu-	sa	es-	qui-	vo”.

**Regra 1001**      elisão      29,7‰      ★

**ber:** /(a)/([aeiouɾɛɔãẽĩõũ])/   **msub:** “2”   **bqsil:** 2   **bpar:** 1  
**bton:** (átona, tônica)   **blim:** (não-inicial e final, inicial)   **mpts:** -3  
**mreaj:** (nada, diminuir)   **mdel:** (nada, excluir)

FIG:2.25.2.2 Do destino feliz, ou infeliz da sua alma,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>antes</i>	du	'des	'ti	'nu	'fe	'lis	'u	'ĩ	'fe	'lis	'da	'su	a	'al	'ma
<i>depois</i>	du	'des	'ti	'nu	'fe	'lis	'u	'ĩ	'fe	'lis	'da	'su	'al	'ma	
	1	2	3	4	5	6	7	8	9	10	11	12	13		
<i>final</i>	du	'des	'ti	'nu	'fe	'li	zwĩ	'fe	'liz	'da	'su	'al	'ma		
<i>final</i>	Do	'des	'ti	'no	'fe	'li-	-z,	ou	in-	'fe	'liz	'da	'sua	'al-	'ma,



## Camada 6 (cont.)

**Regra 1004**      elisão      < 0,1%

**ber:** /(awa)/   **msub:** “wa”   **bqsil:** 2   **bpar:** 1   **bton:** (átona, átona)

**blim:** (não-inicial e final, inicial e final)   **mpts:** 0

**mreaj:** (nada, diminuir)   **mdel:** (nada, excluir)

CAM.10.31.1 “Com toda u’a coxa fora, que em pedaços

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	kõ	'tõ	da	wa	'kõ	ʃa	'fõ	ra	kɪ	ẽ	pe	'da	sus
<i>depois</i>	kõ	'tõ	dwa	'kõ	ʃa	'fõ	ra	kɪ	ẽ	pe	'da	sus	
	1	2	3	4	5	6	7	8	9	10	11		
<i>final</i>	kõ	'tõ	dwa	'kõ	ʃa	'fõ	ra	kjẽ	pe	'da	sus		
<i>final</i>	“Com	to-	da	u’a	co-	xa	fo-	ra,	que	em	pe-	da-	ços

**Regra 1005**      elisão      < 0,1%

**ber:** /(aje)/   **msub:** “je”   **bqsil:** 2   **bpar:** 1   **bton:** (átona, átona)

**blim:** (inicial e final, inicial e não-final)   **mpts:** 2

**mreaj:** (nada, diminuir)   **mdel:** (nada, excluir)

ANJ.24.1.2 É-me grato adstringir-me, na hierarquia

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'ɛ	mɪ	'gra	tu	ads	trĩ	'zir	mɪ	na	je	'rar	'ki	a
<i>depois</i>	'ɛ	mɪ	'gra	tu	ads	trĩ	'zir	mɪ	nje	'rar	'ki	a	
	1	2	3	4	5	6	7	8	9	10	11		
<i>final</i>	'ɛ	mɪ	'gra	twads	trĩ	'zir	mɪ	nje	'rar	'ki	a		
<i>final</i>	É	me	gra	to	ads	trin	gir	me,	na	hie	rar	qui	a

## Camada 7

**Regra 1301**      especial      757,2%      ★

**ber:** /[{ɸsɒdf3ptvsʌɲlsrzmngk}]/   **msub:** “”   **mpts:** 0   **emcong:** sim

ANJ.30.4.3 No pandemônio aterrador do Caos!

	1	2	3	4	5	6	7	8	9	10	11		
<i>antes</i>	nu	pã	de	'mo	nju	a	te	ʒa	'dɔr	du	'kaws		
<i>depois</i>	nu	pã	de	'mo	nju	a	te	ʒa	'dɔr	du	'kaws		
	1	2	3	4	5	6	7	8	9	10			
<i>final</i>	nu	pã	de	'mo	njwa	te	ʒa	'dɔr	du	'kaws			
<i>final</i>	No	pan-	de-	mô-	nio	a-	te-	rra-	dor	do	Caos!		

## Camada 8

**Regra 501** especial 485,3‰ ★

**ber:** /((s)([aãeēēĩĩõõçõũjwbdgʒlʌmŋɣɲɸ])/  
**msub:** “z\2” **bqsil:** 2 **bpar:** 1  
**mpts:** -1 **ebcong:** sim

SIL:5.167.3.2 Feito d'aromas, de doçuras feito

	1	2	3	4	5	6	7	8	9	10	11
<i>antes</i>	'	f	e	j	t	u	:	d	a	:	'
	r	o	:	m	a	s	:	d	i	:	'
	d	o	:	s	u	:	r	a	s	:	'
	f	e	j	:	t	u	:	d	a	:	'
<i>depois</i>	'	f	e	j	t	u	:	d	a	:	'
	r	o	:	m	a	z	:	d	i	:	'
	d	o	:	s	u	:	r	a	s	:	'
	f	e	j	:	t	u	:	d	a	:	'
<i>final</i>	'	f	e	j	:	t	u	:	d	a	:
	r	o	:	m	a	z	:	d	i	:	'
	d	o	:	s	u	:	r	a	s	:	'
<i>final</i>	F	e	i	:	t	o	:	d	a	:	'
	r	o	:	m	a	s	:	d	e	:	'
	d	o	:	ç	u	:	r	a	s	:	'
	F	e	i	:	t	o	:	d	a	:	'

**Regra 502** especial 254,7‰ ★

**ber:** /([zr])([aãeēēĩĩõõçõũjw])/  
**msub:** “\1\2” **bqsil:** 2 **bpar:** 1 **mpts:** -1  
**mreaj:** (diminuir, aumentar) **ebcong:** sim

COS.101.12.12 Que sustentar a sólida aliança

	1	2	3	4	5	6	7	8	9	10	11
<i>antes</i>	k	i	:	s	u	s	:	t	ẽ	:	'
	t	a	:	r	a	:	a	:	s	o	:
	s	o	:	l	i	:	d	a	:	l	i
	ã	:	s	a	:	'					
<i>depois</i>	k	i	:	s	u	s	:	t	ẽ	:	'
	t	a	:	r	a	:	a	:	s	o	:
	s	o	:	l	i	:	d	a	:	l	i
	ã	:	s	a	:	'					
<i>final</i>	k	i	:	s	u	s	:	t	ẽ	:	'
	t	a	:	r	a	:	a	:	s	o	:
	s	o	:	l	i	:	d	a	:	l	i
	ã	:	s	a	:	'					
<i>final</i>	Q	u	e	:	s	u	s	:	t	e	n
	t	a	:	r	a	:	s	o	:	l	i
	d	a	:	a	:	l	i	:	a	n	ç
	a	:									

## Camada 9

**Regra 2102** crase 62,7‰ ★

**ber:** /([aɪ])(\2)/ **msub:** “\2” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona)  
**mpts:** -2 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

MAT:5.48.6.3 Quando o todo é suor, e porcaria,

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'	k	w	ã	:	d	u	:	u	:	'	t	o
	d	u	:	'	ɛ	:	s	u	:	'	ɔ	:	r
	r	i	:	p	o	:	r	k	a	:	'	r	i
	a	:	'										
<i>depois</i>	'	k	w	ã	:	d	u	:	'	t	o	:	d
	u	:	'	ɛ	:	s	u	:	'	ɔ	:	r	i
	p	o	:	r	k	a	:	'	r	i	:	a	:
	'												
<i>final</i>	'	k	w	ã	:	d	u	:	'	t	o	:	d
	w	ɛ	:	s	u	:	'	ɔ	:	r	i	:	p
	o	:	r	k	a	:	'	r	i	:	a	:	'
<i>final</i>	Q	u	a	n	-	d	o	-	t	o	-	d	o
	é	-	s	u	-	o	-	r	,e	-	p	o	r
	-	c	a	-	r	i	-	a	,				

## Camada 9 (cont.)

**Regra 2103** crase 27,2‰ ★

**ber:** /(<ʊ|ɪ|i|ɪ|ɪ|ʊ|ʊ)/ **msub:** “X” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona)

**mpts:** -2 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 1 com mapa [uiĩui] → [uiĩui])

CAM.8.96.7 Pode o vil interesse e sede imiga

	1	2	3	4	5	6	7	8	9	10	11	12	13													
<i>antes</i>	p	ɔ	d	r	u	ˈ	v	i	l	t	e	r	ɛ	s	i	ˈ	s	e	d	i	ˈ	m	i	ˈ	g	a
<i>depois</i>	p	ɔ	d	r	u	ˈ	v	i	l	t	e	r	ɛ	s	i	ˈ	s	e	d	i	ˈ	m	i	ˈ	g	a

	1	2	3	4	5	6	7	8	9	10	11															
<i>final</i>	p	ɔ	d	j	u	ˈ	v	i	l	t	e	r	ɛ	s	i	ˈ	s	e	d	i	ˈ	m	i	ˈ	g	a
<i>final</i>	Po	-	de	o	vil	-	in	-	te	-	re	-	sse	e	-	se	-	de	i	-	mi	-	ga			

**Regra 2111** crase 7,1‰ ★

**ber:** /(<a|ʊ|ɪ|ɪ)/ **msub:** “X” **bqsil:** 2 **bpar:** 1 **bton:** (átona, tônica)

**mpts:** -1 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 1 com mapa [aui] → [aui])

MAT:29.113.1.7 que isso a todos nos espanta,

	1	2	3	4	5	6	7	8	9	10																
<i>antes</i>	k	i	ˈ	s	u	ˈ	a	ˈ	t	o	d	u	z	ˈ	n	u	ˈ	z	e	s	ˈ	p	ã	ˈ	t	a
<i>depois</i>	k	i	ˈ	s	u	ˈ	a	ˈ	t	o	d	u	z	ˈ	n	u	ˈ	z	e	s	ˈ	p	ã	ˈ	t	a

	1	2	3	4	5	6	7	8																	
<i>final</i>	k	i	ˈ	s	w	a	ˈ	t	o	d	u	z	ˈ	n	u	ˈ	z	e	s	ˈ	p	ã	ˈ	t	a
<i>final</i>	que	i	-	sso	a	-	to	-	dos	-	no	-	s	es	-	pan	-	ta	,						

**Regra 2106** crase 1,8‰

**ber:** /(<a|e|e)/ **msub:** “X” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona) **mpts:** -2

**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [ae] → [ae])

MAT:5.30.3.9 se virá a queixar, senão

	1	2	3	4	5	6	7	8																	
<i>antes</i>	s	i	ˈ	v	i	ˈ	r	a	ˈ	a	ˈ	k	e	j	ˈ	ʃ	a	r	ˈ	s	e	ˈ	n	ã	w
<i>depois</i>	s	i	ˈ	v	i	ˈ	r	a	ˈ	k	e	j	ˈ	ʃ	a	r	ˈ	s	e	ˈ	n	ã	w	ˈ	

	1	2	3	4	5	6	7																
<i>final</i>	s	i	ˈ	v	i	ˈ	r	a	ˈ	k	e	j	ˈ	ʃ	a	r	ˈ	s	e	ˈ	n	ã	w
<i>final</i>	se	-	vi	-	r	a	-	quei	-	xar	-	se	-	n	ã	o							

## Camada 9 (cont.)

**Regra 2105** crase 0,2‰

**ber:** /(i|ii|uɔ)/ **msub:** “X” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)

**mpts:** -1 **mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [iu] → [ju])

CAM.10.23.3 Em ti e nele veremos altos peitos

	1	2	3	4	5	6	7	8	9	10	11	12																
<i>antes</i>	ẽ	ˈ	ti	ˌ	ne	ˌ	li	ˌ	ve	ˌ	re	ˌ	mu	ˌ	ˈ	z	al	ˌ	tus	ˌ	pe	ˌ	j	ˌ	tus			
<i>depois</i>	ẽ	ˈ	ti	ˌ	ne	ˌ	li	ˌ	ve	ˌ	re	ˌ	mu	ˌ	ˈ	z	al	ˌ	tus	ˌ	pe	ˌ	j	ˌ	tus			
<i>final</i>	ẽ	ˈ	ti	ˌ	ne	ˌ	li	ˌ	ve	ˌ	re	ˌ	mu	ˌ	ˈ	z	al	ˌ	tus	ˌ	pe	ˌ	j	ˌ	tus			
<i>final</i>	Em	ˌ	ti	ˌ	e	ˌ	ne	ˌ	le	ˌ	ve	ˌ	re	ˌ	mo	ˌ	s	al	ˌ	to	ˌ	s	pe	ˌ	i	ˌ	to	s

**Regra 2112** crase 0,2‰

**ber:** /(j|wɔ)/ **msub:** “X” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona) **mpts:** 7

**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

**mmap:** (em 0 inserir 0 com mapa [jw] → [jw])

CAM.2.113.1 “Queimou o sagrado templo de Diana,

	1	2	3	4	5	6	7	8	9	10	11	12																	
<i>antes</i>	kej	ˌ	ˈ	mow	ˌ	u	ˌ	sa	ˌ	ˈ	gra	ˌ	du	ˌ	ˈ	tē	ˌ	plu	ˌ	d	i	ˌ	d	i	ˌ	ˈ	ã	ˌ	na
<i>depois</i>	kej	ˌ	ˈ	mow	ˌ	sa	ˌ	ˈ	gra	ˌ	du	ˌ	ˈ	tē	ˌ	plu	ˌ	d	i	ˌ	d	i	ˌ	ˈ	ã	ˌ	na		
<i>final</i>	kej	ˌ	ˈ	mow	ˌ	sa	ˌ	ˈ	gra	ˌ	du	ˌ	ˈ	tē	ˌ	plu	ˌ	d	i	ˌ	d	i	ˌ	ˈ	ã	ˌ	na		
<i>final</i>	“Quei-	ˌ	mou	ˌ	o	ˌ	sa-	ˌ	gra-	ˌ	do	ˌ	tem-	ˌ	plo	ˌ	de	ˌ	Di-	ˌ	a-	ˌ	na,						

**Regra 2117** crase 0,1‰

**ber:** /(ɔɔu)/ **msub:** “u” **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, indiferente)

**mpts:** -1 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

ANJ.63.2.4 De todo o ubiquitário Movimento!

	1	2	3	4	5	6	7	8	9	10	11	12	13																		
<i>antes</i>	d	ˌ	ˈ	to	ˌ	du	ˌ	u	ˌ	b	i	ˌ	k	i	ˌ	ˈ	ta	ˌ	r	j	u	ˌ	mo	ˌ	vi	ˌ	ˈ	mē	ˌ	t	u
<i>depois</i>	d	ˌ	ˈ	to	ˌ	du	ˌ	b	i	ˌ	k	i	ˌ	ˈ	ta	ˌ	r	j	u	ˌ	mo	ˌ	vi	ˌ	ˈ	mē	ˌ	t	u		
<i>final</i>	d	ˌ	ˈ	to	ˌ	du	ˌ	b	i	ˌ	k	i	ˌ	ˈ	ta	ˌ	r	j	u	ˌ	mo	ˌ	vi	ˌ	ˈ	mē	ˌ	t	u		
<i>final</i>	De	ˌ	to	ˌ	do	ˌ	o	u	ˌ	bi	ˌ	qui	ˌ	tá	ˌ	rio	ˌ	Mo	ˌ	vi	ˌ	men	ˌ	to							

## Camada 9 (cont.)

**Regra 2104** crase < 0,1‰

**ber:** /⟨iĩ⟩/ **msub:** “ĩ” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona) **mpts:** -2  
**mdel:** (nada, excluir)

MAG.9.10.4 Até aqui indomáveis;

	1	2	3	4	5	6	7	8
<i>antes</i>	a	ˈtɛ	a	ˈki	ĩ	do	ˈma	ˈvijs
<i>depois</i>	a	ˈtɛ	a	ˈkiĩ	do	ˈma	ˈvijs	
	1	2	3	4	5	6	7	
<i>final</i>	a	ˈtɛ	a	ˈkiĩ	do	ˈma	ˈvijs	
<i>final</i>	A	té	a	qui	in	do	má	veis;

**Regra 2110** crase < 0,1‰

**ber:** /⟨iĩ⟩/ **msub:** “ĩ” **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, átona)  
**mpts:** -2 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

CAM.10.72.2 Vencedor forte e intrépido, ao possante

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	vê	ˈse	ˈdɔr	ˈfɔr	tĩ	ĩ	ˈtrɛ	pi	du	ˈaw	po	ˈsã	tĩ	
<i>depois</i>	vê	ˈse	ˈdɔr	ˈfɔr	tĩ	ˈtrɛ	pi	du	ˈaw	po	ˈsã	tĩ		
	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	vê	ˈse	ˈdɔr	ˈfɔr	tĩ	ˈtrɛ	pi	ˈdwaw	po	ˈsã	tĩ			
<i>final</i>	Ven	-ce	-dor	for	-te	e	in	-tré	-pi	-do	,ao	po	-ssan	-te

**Regra 2113** crase < 0,1‰

**ber:** /⟨i⟩/ **msub:** “j” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona) **mpts:** 0  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

ANJ.183.4.1 E em meio às refrações verdes e hialinas,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ĩ	ˈẽ	ˈmej	ˈu	ˈaz	ˈɐ	ˈfra	ˈsɔjz	ˈvɛr	ˈdĩ	ˈzi	ˈja	ˈli	ˈnas
<i>depois</i>	ĩ	ˈẽ	ˈmej	ˈu	ˈaz	ˈɐ	ˈfra	ˈsɔjz	ˈvɛr	ˈdĩ	ˈzja	ˈli	ˈnas	
	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	jẽ	ˈmej	ˈwaz	ˈɐ	ˈfra	ˈsɔjz	ˈvɛr	ˈdĩ	ˈzja	ˈli	ˈnas			
<i>final</i>	E	em	mei	-o	às	re	-fra	-ções	ver	-de	-s	e	hial	-inas,



### Camada 9 (cont.)

**Regra 2101**      crase       $< 0,1\%$

ber: /(([aɪʊ])(2\2))/ msub: “2” bqsil: 3 bpar: 1

**btón:** (átóna, átóna, átóna) **mpts:** -3 **mreaj:** (nada, diminuir, disminuir)

**mdel:** (nada, exclure, exclure)

ANJ.64.2.2 Vence o granito, deprimindo-o... O espanto

antes 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15  
*antes* vê: sɪ u gra: ni tu de pri mĩ du u u es pã tu  
 depois 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15  
*depois* vê: sɪ u gra: ni tu de pri mĩ du es pã tu

1 2 3 4 5 6 7 8 9 10 11  
*final* vê sju gra: ni tu de pri mĩ dwes pã tu  
*final* Ven- ce o gra- ni- to, de- pri- mín- do-o... O es- pan- to

**Regra 2116** crase < 0,1‰

**ber:** /(III)/ **msub:** “i” **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, átona)

**mpts:** -1 **mreaaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

CAM.5.80.1 "Mas não foi, da esperança grande e imensa

antes maz nāw foj da es pe rā sa grā dī i i mē sa  
depois maz nāw foj da es pe rā sa grā di mē sa

final maz nāw foj da es pe rā sa grā di mē sa  
final "Mas não foi, da es pe ran ça gran de e i men sa

**Regra 2114**      crase      < 0,1‰

ber: /ʔi/ msub: “j” bqsil: 2 bpar: 1 bton: (átona, tônica) mpts: 2

**mreaj:** (nada, diminuir)    **mdel:** (nada, excluir)

ANJ.12.77.1 A água arbitrária que hiulcos caules grossos

antes 1 2 3 4 5 6 7 8 9 10 11 12  
 'a gwar bi tra rja kɪ 'jul kus kaw lɪz grɔ sus  
 depois 'a gwar bi tra rja kjul kus kaw lɪz grɔ sus

final 1 2 3 4 5 6 7 8 9 10 11  
 'a gwar bi tra rja 'kjul kus kaw lɪz grɔ sus  
 final A â gua ar bi trá ria que hiul cos cau les gro ssos

## Camada 9 (cont.)

**Regra 2115** crase < 0,1‰

**ber:** /⟨ɪj⟩/ **msub:** “j” **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, átona)  
**mpts:** 3 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

ANJ.257.1.2 Das que, da excele e hierática hierarquia,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	das	kr	da	e	se	lɪ	ɪ	je	ra	ti	kje	rar	ki	a
<i>depois</i>	das	kr	da	e	se	lje	ra	ti	kje	rar	ki	a		

	1	2	3	4	5	6	7	8	9	10	11
<i>final</i>	das	kr	dag	se	lje	ra	ti	kje	rar	ki	a
<i>final</i>	Das	que,	da	e-xce-	le	e	hie-rá-	ti-	ca	hie-rar-	qui-a,

**Regra 2107** crase < 0,1‰

**ber:** /⟨aaa⟩/ **msub:** “a” **bqsil:** 3 **bpar:** 1 **bton:** (tônica, átona, átona)  
**mpts:** -2 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

CAM.7.28.3 Como se longa fora já a amizade,

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'ko	mu	si	'lõ	ga	'fõ	ra	'za	a	a	'mi	'za	dɪ
<i>depois</i>	'ko	mu	si	'lõ	ga	'fõ	ra	'za	'mi	'za	dɪ		

	1	2	3	4	5	6	7	8	9	10	11
<i>final</i>	'ko	mu	si	'lõ	ga	'fõ	ra	'za	'mi	'za	dɪ
<i>final</i>	Co-	mo-	se	lon-	ga-	fo-	ra	já	a	-mi-	za-de,

**Regra 2108** crase < 0,1‰

**ber:** /⟨aaa⟩/ **msub:** “a” **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, tônica)  
**mpts:** 0 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

ANJ.33.1.2 E atira-a à arca ancestral dos palimpsestos...

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ɪ	a	'ti	ra	a	'ar	kā	ses	'tral	dus	pa	lĩp	'ses	tus
<i>depois</i>	ɪ	a	'ti	rar	kā	ses	'tral	dus	pa	lĩp	'ses	tus		

	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	ja	'ti	'rar	kā	ses	'tral	dus	pa	lĩp	'ses	tus			
<i>final</i>	E	a-	ti-	ra-a	à	ar-	ca	an-	ces-	tral	dos	pa-	limp-	ses-tos...

## Camada 9 (cont.)

**Regra 2109** crase < 0,1‰

**ber:** /(aaa)/ **msub:** “a” **bqsil:** 3 **bpar:** 1 **bton:** (átona, tônica, átona)

**mpts:** 0 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

CAM.10.140.4 Descobri-la-á a primeira vossa frota.

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	des	ko	'bri	la	'a	a	'pri	'mej	ra	'vɔ	sa	'frɔ	ta
<i>depois</i>	des	ko	'bri	'la	'pri	'mej	ra	'vɔ	sa	'frɔ	ta		
	1	2	3	4	5	6	7	8	9	10	11		
<i>final</i>	des	ko	'bri	'la	'pri	'mej	ra	'vɔ	sa	'frɔ	ta		
<i>final</i>	Des	-co	-bri	-la	-á	a	'pri	-mei	-ra	-vo	-ssa	-fro	-ta.

## Camada 10

**Regra 2501** sinalefa 398,0‰ ★

**ber:** /([iɪ])([aeouãẽõũɛɔɯ])/ **msub:** “j2” **bqsil:** 2 **bpar:** 1

**bton:** (átona, indiferente) **mpts:** -3 **mdel:** (nada, excluir)

CAM.7.58.4 Que a seu costume estava ruminando.

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	kja	'sew	kos	'tu	mɪ	es	'ta	va	'ɸu	mi	'nã	du
<i>depois</i>	kja	'sew	kos	'tu	mjes	'ta	va	'ɸu	mi	'nã	du	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	kja	'sew	kos	'tu	mjes	'ta	va	'ɸu	mi	'nã	du	
<i>final</i>	Que	a	seu	cos	-tu	-me	es	-ta	-va	-ru	-mi	-nan-do.

**Regra 2502** sinalefa 226,1‰ ★

**ber:** /([ʊ])([aeioãẽĩõũɛɔɪ])/ **msub:** “w2” **bqsil:** 2 **bpar:** 1

**bton:** (átona, indiferente) **mpts:** -2 **mdel:** (nada, excluir)

COS.143.2.5 Inspiras de teu raio esforço ativo?

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	ĩs	'pi	raz	dɪ	'tew	'ɸaj	wes	'fɔr	su	'a	'ti	vu
<i>depois</i>	ĩs	'pi	raz	dɪ	'tew	'ɸaj	wes	'fɔr	swa	'ti	vu	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	ĩs	'pi	raz	dɪ	'tew	'ɸaj	wes	'fɔr	swa	'ti	vu	
<i>final</i>	Ins	-pi	-ras	de	teu	rai	-o	es	-for	-ço	a	-ti-vo?

## Camada 10 (cont.)

**Regra 2503** sinalefa 21,8% ★

**ber:** /([əuɪiĩēōāoeū])/ **msub:** “aX” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona)

**mpts:** -1 **mdel:** (nada, excluir)

**mmap:** (em 1 inserir 1 com mapa [əuɪiĩēōāoeū] → [əwwjjiĩēōāoeū])

MAT:32.1.1.54 o qual, levado da inveja

	1	2	3	4	5	6	7	8	9	
<i>antes</i>	u	kw	al	le	va	du	da	ĩ	vẽ	za
<i>depois</i>	u	kw	al	le	va	du	daĩ	vẽ	za	

	1	2	3	4	5	6	7	8	
<i>final</i>	u	kw	al	le	va	du	daĩ	vẽ	za
<i>final</i>	o	qual	le	va	do	da	in	ve	ja

**Regra 2517** sinalefa 8,3% ★

**ber:** /([ɪʊ]a[ãẽũĩōēiου])/ **msub:** “XaY” **bqsil:** 3 **bpar:** 1

**bton:** (átona, átona, átona) **mpts:** 0 **mdel:** (nada, excluir, excluir)

**mmap:** (em 0 inserir 0 com mapa [ɪʊ] → [jw], em 2 inserir 2 com mapa [ãẽũĩōējɔww])

ANJ.67.24.6 Com a terra do sepulcro a encher-lhe os olhos!

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	kwa	te	xa	du	se	pul	kru	a	ẽ	ʃer	ʃju	zo	ʃus
<i>depois</i>	kwa	te	xa	du	se	pul	krwaẽ	ʃer	ʃju	zo	ʃus		

	1	2	3	4	5	6	7	8	9	10	11					
<i>final</i>	kwa	te	xa	du	se	pul	krwaẽ	ʃer	ʃju	zo	ʃus					
<i>final</i>	Com	a	te	rra	do	se	pul	cro	a	en	cher	lhe	o	s	o	lhos

**Regra 2516** sinalefa 8,0% ★

**ber:** /((ɪʊ)ɪ) ([ãẽĩūaeiouɪ])/ **msub:** “XY3” **bqsil:** 3 **bpar:** 1

**bton:** (átona, átona, átona) **mpts:** 1 **mdel:** (nada, excluir, excluir)

**mmap:** (em 0 inserir 0 com mapa [ɪʊ] → [jw], em 1 inserir 1 com mapa [ɪʊ] → [jw])

FIG:5.19.1.2 De branco, e assim coroadas de alvas flores

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	dĩ	brã	ku	ĩ	a	sĩ	ko	rwa	da	djal	vas	fl	ris
<i>depois</i>	dĩ	brã	kwja	sĩ	ko	rwa	da	djal	vas	fl	ris		

	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	dĩ	brã	kwja	sĩ	ko	rwa	da	djal	vas	fl	ris			
<i>final</i>	De	bran	co	e	a	ssim	co	roa	da	de	al	vas	flo	res

### Camada 10 (cont.)

**Regra 2506** sinalefa 2,8‰

**ber:** /([æɔeoã])[ʊuĩĩĩ]/ **msub:** “2X” **bqsil:** 2 **bpar:** 1  
**bton:** (tônica, átona) **mpts:** -2 **mdel:** (nada, excluir)  
**mmap:** (em 1 inserir 1 com mapa [ʊuĩĩĩ] → [wwũjĩĩ])

MAT:5.30.4.6 nem menos está o seu mal

	1	2	3	4	5	6	7	8
<i>antes</i>	'nē	'me	nu	zes	'ta	u	'sew	'mal
<i>depois</i>	'nē	'me	nu	zes	'taw	'sew	'mal	

	1	2	3	4	5	6	7
<i>final</i>	'nē	'me	nu	zes	'taw	'sew	'mal
<i>final</i>	nem	me	no	s es	tá o	seu	mal

**Regra 2505** sinalefa 1,9‰ ★

**ber:** /(ĩĩ)/ **msub:** “jĩ” **bqsil:** 2 **bpar:** 1 **bton:** (átona, tônica) **mpts:** -2  
**mdel:** (nada, excluir)

PIN.34.23.2 Mas a partir a noite insta, assomando:

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	ma	za	par	'ti	ra	'noj	ti	'ts	ta	so	'mã	du
<i>depois</i>	ma	za	par	'ti	ra	'noj	'tjĩs	ta	so	'mã	du	

	1	2	3	4	5	6	7	8	9	10	11
<i>final</i>	ma	za	par	'ti	ra	'noj	'tjĩs	ta	so	'mã	du
<i>final</i>	Ma	s a	par	ti	r a	noi	te ins	ta,	a-	sso	man-do:

**Regra 2504** sinalefa 1,8‰

**ber:** /(a([oõɔeɛãĩũ])/ **msub:** “aX” **bqsil:** 2 **bpar:** 1 **bton:** (átona, tônica)  
**mpts:** 3 **mdel:** (nada, excluir)  
**mmap:** (em 1 inserir 1 com mapa [oõɔeɛãĩũ] → [oõɔɔeɛãjĩw])

MAT:2.29.73.6 a este barro nada fino

	1	2	3	4	5	6	7	8	9
<i>antes</i>	a	'es	ti	'ba	ɥu	'na	da	'fi	nu
<i>depois</i>	'aɛs	ti	'ba	ɥu	'na	da	'fi	nu	

	1	2	3	4	5	6	7	8
<i>final</i>	'aɛs	ti	'ba	ɥu	'na	da	'fi	nu
<i>final</i>	a es	te	ba	rro	na	da	fi	no



# Camada 10 (cont.)

**Regra 2507** sinalefa 0,9‰

**ber:** /([æɔã])[oõeē]/ **msub:** “2X” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)  
**mpts:** 0 **mdel:** (nada, excluir)  
**mmap:** (em 1 inserir 1 com mapa [oõeē] → [oõeē])

MAT:4.35.1.134 da popa de Nau nova

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	da	ˈpɔ	ˈpa	ˈdɪ	ˈnaw	ˈnɔ	ˈves	ˈta	ẽ	ˈta	ˈʎa	ˈda
<i>depois</i>	da	ˈpɔ	ˈpa	ˈdɪ	ˈnaw	ˈnɔ	ˈves	ˈtaẽ	ˈta	ˈʎa	ˈda	
	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	da	ˈpɔ	ˈpa	ˈdɪ	ˈnaw	ˈnɔ	ˈves	ˈtaẽ	ˈta	ˈʎa	ˈda	
<i>final</i>	da	ˈpo	ˈpa	ˈde	ˈNau	ˈno	ˈva	ˈes	ˈtá	ˈen	ˈta	ˈlha
												ˈda

**Regra 2508** sinalefa 0,8‰

**ber:** /([i[uɔũ]]u[ĩĩr])/ **msub:** “XY” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)  
**mpts:** -1 **mdel:** (nada, excluir) **mmap:** (em 0 inserir 0 com mapa [iu] → [iu], em 1 inserir 1 com mapa [uũĩĩr] → [wũĩĩĩĩ])

MAT:29.132.1.1 Corre por aqui uma voz,

	1	2	3	4	5	6	7	8
<i>antes</i>	ˈkɔ	ˈʁɪ	ˈpu	ˈra	ˈki	u	ˈma	ˈvɔs
<i>depois</i>	ˈkɔ	ˈʁɪ	ˈpu	ˈra	ˈkiw	ˈma	ˈvɔs	
	1	2	3	4	5	6	7	
<i>final</i>	ˈkɔ	ˈʁɪ	ˈpu	ˈra	ˈkiw	ˈma	ˈvɔs	
<i>final</i>	Co	ˈrre	ˈpo	ˈr	a	ˈqui	u	ˈma
								ˈvoz,

**Regra 2520** sinalefa 0,6‰

**ber:** /([ɪʊ]a[ãɔɔẽɔĩu])/ **msub:** “XaY” **bqsil:** 3 **bpar:** 1  
**bton:** (átona, átona, tônica) **mpts:** 3 **mdel:** (nada, excluir, excluir)  
**mmap:** (em 0 inserir 0 com mapa [ɪʊ] → [jw], em 2 inserir 2 com mapa [ãɔɔẽɔĩu] → [ãɔɔẽɔĩjw])

PIN.79.49.1 E comigo a esta santa companhia.”

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	ɪ	ˈko	ˈmi	ˈgu	a	ˈɛs	ˈta	ˈsã	ˈta	ˈkõ	ˈpa	ˈɲi	ˈa
<i>depois</i>	ɪ	ˈko	ˈmi	ˈgwaɣs	ˈta	ˈsã	ˈta	ˈkõ	ˈpa	ˈɲi	ˈa		
	1	2	3	4	5	6	7	8	9	10	11		
<i>final</i>	ɪ	ˈko	ˈmi	ˈgwaɣs	ˈta	ˈsã	ˈta	ˈkõ	ˈpa	ˈɲi	ˈa		
<i>final</i>	E	ˈco	ˈmi	ˈgo	a	ˈes	ˈta	ˈsan	ˈta	ˈcom	ˈpa	ˈnhi	ˈa

## Camada 10 (cont.)

**Regra 2512** sinalefa 0,4%

**ber:** /([aẽo]))/ **msub:** “j2” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona)  
**mpts:** 8 **mdel:** (nada, excluir) **mton:** (tornar átona, nada)

CAM.9.88.1 Assi a formosa e a forte companhia

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	a	si	a	for	mo	zja	fôr	ti	kõ	pa	ni	a
<i>depois</i>	a	sja	for	mo	zja	fôr	ti	kõ	pa	ni	a	

	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	a	sja	for	mo	zja	fôr	ti	kõ	pa	ni	a			
<i>final</i>	A	ssi	a	for	mo	sa	e	a	for	te	com	pa	nhi	a

**Regra 2514** sinalefa 0,3%

**ber:** /((([æɔ])([æeɔõãĩ]))/ **msub:** “\2X” **bqsil:** 2 **bpar:** 1  
**bton:** (tônica, tônica) **mpts:** 3 **mdel:** (nada, excluir)  
**mmap:** (em 1 inserir 1 com mapa [æeɔõãĩ] → [æ̃ẽɔ̃õ̃ã̃ĩ̃])

FIG:5.84.2.2 Sonhos possas levar, ó alma amorosa,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	'so	jus	'po	saz	le	'va	'rɔ	'al	ma	mo	'rɔ	za
<i>depois</i>	'so	jus	'po	saz	le	'va	rɔal	ma	mo	'rɔ	za	

	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	'so	jus	'po	saz	le	'va	rɔal	ma	mo	'rɔ	za			
<i>final</i>	So	-nhos	po	-ssas	le	-va	-r,	ó	al	-ma	a	-mo	-ro	-sa,

**Regra 2515** sinalefa 0,2%

**ber:** /((([i])([ɔẽõ]))/ **msub:** “j3” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, tônica)  
**mpts:** -1 **mdel:** (nada, excluir)

MAT:5.18.11.9 porém fique aqui entre nós.

	1	2	3	4	5	6	7	8
<i>antes</i>	po	rẽ	'fi	kja	'ki	ẽ	tri	'nos
<i>depois</i>	po	rẽ	'fi	kja	'kjẽ	tri	'nos	

	1	2	3	4	5	6	7		
<i>final</i>	po	rẽ	'fi	kja	'kjẽ	tri	'nos		
<i>final</i>	po	-rém	fi	-que	a	-qui	en	-tre	nós.



## Camada 10 (cont.)

**Regra 2551** sinalefa 0,1‰

**ber:** /(awī)/ **msub:** “wī” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona) **mpts:** 5  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

PIN.70.40.2 Deu de Suábia ao Imperador segundo

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	dew	di	su	a	bjaw	ī	pe	ra	dor	se	gū	du
<i>depois</i>	dew	di	su	a	bjwī	pe	ra	dor	se	gū	du	

	1	2	3	4	5	6	7	8	9	10	11		
<i>final</i>	dew	di	su	a	bjwī	pe	ra	dor	se	gū	du		
<i>final</i>	Deu	de	Su-	á-	bia	ao	Im-	pe-	ra-	dor	se-	gun-	do

**Regra 2510** sinalefa < 0,1‰

**ber:** /(e[aãoô][o[aãēē]])/ **msub:** “XY” **bqsil:** 2 **bpar:** 1  
**bton:** (tônica, átona) **mpts:** 3 **mdel:** (nada, excluir) **mmap:** (em 0 inserir 0 com mapa [eo] → [eo], em 1 inserir 1 com mapa [aãoōēē] → [aãōōēē])

MAT:5.47.2.5 como não vê a concorrência

	1	2	3	4	5	6	7	8	9
<i>antes</i>	ko	mu	nāw	ve	a	kō	ko	ʁē	sja
<i>depois</i>	ko	mu	nāw	veq	kō	ko	ʁē	sja	

	1	2	3	4	5	6	7	8	
<i>final</i>	ko	mu	nāw	veq	kō	ko	ʁē	sja	
<i>final</i>	co-	mo	não	vê	a	con-	co-	rrên-	cia

**Regra 2553** sinalefa < 0,1‰

**ber:** /(awa)/ **msub:** “wa” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona) **mpts:** 5  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

FIG:5.4.2.4 Ou gotas de água ao afago da impiedade?

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	ow	gō	taz	dja	gwaw	a	fa	gu	daĩ	pje	da	dĩ
<i>depois</i>	ow	gō	taz	dja	gwwa	fa	gu	daĩ	pje	da	dĩ	

	1	2	3	4	5	6	7	8	9	10	11				
<i>final</i>	ow	gō	taz	dja	gwwa	fa	gu	daĩ	pje	da	dĩ				
<i>final</i>	Ou	go-	tas	de	á-	gua	ao	a-	fa-	go	da	im-	pie-	da-	de?

## Camada 10 (cont.)

**Regra 2554** sinalefa < 0,1‰

**ber:** /{(owa)/ **msub:** “oã” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona) **mpts:** 5  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

FIG:3.13.3.2 Não quis viver; voltou ao áureo caminho,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	'nāw	'kiz	vi	'ver	vol	'tow	'aw	'aw	rju	'ka	'mi	jnu
<i>depois</i>	'nāw	'kiz	vi	'ver	vol	toq̃w	'aw	rju	'ka	'mi	jnu	

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	'nāw	'kiz	vi	'ver	vol	toq̃w	'aw	rju	'ka	'mi	jnu	
<i>final</i>	Não	quis	vi	'ver	vol	-tou	ao	áu	-reo	-ca	-mi	-nho,

**Regra 2534** sinalefa < 0,1‰

**ber:** /{(ea[eio])/ **msub:** “εX” **bqsil:** 3 **bpar:** 1 **bton:** (tônica, átona, átona)  
**mpts:** 5 **mreaj:** (nada, diminuir, nada) **mdel:** (nada, excluir, excluir)  
**mmap:** (em 1 inserir 2 com mapa [eio] → [εjɔ])

DUR.10.43.5 Eis aqui! esta é a imagem veneranda,

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'ej	za	'ki	'es	'te	a	i	'ma	zẽ	ve	'ne	'rã	da
<i>depois</i>	'ej	za	'ki	'es	'tej	'ma	zẽ	ve	'ne	'rã	da		

	1	2	3	4	5	6	7	8	9	10	11				
<i>final</i>	'ej	za	'ki	'es	'tej	'ma	zẽ	ve	'ne	'rã	da				
<i>final</i>	Ei	-s	a	-quil	'es	-ta	é	a	i	-ma	-gem	-ve	-ne	-ran	-da,

**Regra 2528** sinalefa < 0,1‰

**ber:** /{(Iwa)/ **msub:** “jwa” **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona) **mpts:** 1  
**mdel:** (nada, excluir)

CAM.3.141.8 Depois que u'a moça vil na Apúlia viste.

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	de	'pojs	kɪ	wa	'mɔ	sa	'vil	na	'pu	lja	'vis	tɪ
<i>depois</i>	de	'pojs	kjwa	'mɔ	sa	'vil	na	'pu	lja	'vis	tɪ	

	1	2	3	4	5	6	7	8	9	10	11		
<i>final</i>	de	'pojs	kjwa	'mɔ	sa	'vil	na	'pu	lja	'vis	tɪ		
<i>final</i>	De	-pois	que	u'a	-mo	-ça	vil	na	A	-pú	-lia	-vis	-te.

# Camada 10 (cont.)

**Regra 2536** sinalefa < 0,1‰

**ber:** /(\u\u([oĩ]))/ **msub:** "jw\2" **bqsil:** 4 **bpar:** 1

**bton:** (átona, átona, átona, átona) **mpts:** 5

**mreaj:** (nada, diminuir, nada, nada) **mdel:** (nada, excluir, excluir, excluir)

ANJ.25.12.3 O hidrogênio e o oxigênio que tu choras

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
<i>antes</i>	wi	dro	ze	nju	i	u	o	fi	ze	nju	ki	tu	ʃo	ras	
<i>depois</i>	wi	dro	ze	njjwo	fi	ze	nju	ki	tu	ʃo	ras				
<i>final</i>	wi	dro	ze	njjwo	fi	ze	nju	ki	tu	ʃo	ras				
<i>final</i>	O	hi	dro	gê	nio	e	o	o	xi	gê	nio	que	tu	cho	ras

**Regra 2552** sinalefa < 0,1‰

**ber:** /(\uwa)/ **msub:** "wa" **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona) **mpts:** 5

**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

FIG:5.67.2.2 Surgem-se-me, assim, no peito, uma por uma...

	1	2	3	4	5	6	7	8	9	10	11	12	
<i>antes</i>	'sur	zẽ	sɪ	mja	'sɪ	nu	'pej	tu	wa	pu	'ru	ma	
<i>depois</i>	'sur	zẽ	sɪ	mja	'sɪ	nu	'pej	twa	pu	'ru	ma		
<i>final</i>	'sur	zẽ	sɪ	mja	'sɪ	nu	'pej	twa	pu	'ru	ma		
<i>final</i>	Sur	gem	se	me	a	ssim	no	pei	to	uma	po	r u	ma...

**Regra 2518** sinalefa < 0,1‰

**ber:** /(\uēa)/ **msub:** "wĩa" **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, átona)

**mpts:** 5 **mdel:** (nada, excluir, excluir)

MAT:7.9.1.10 em sangue, como em apelido.

	1	2	3	4	5	6	7	8	9	10
<i>antes</i>	ẽ	'sã	gɪ	'ko	mu	ẽ	a	'pe	'li	du
<i>depois</i>	ẽ	'sã	gɪ	'ko	mwĩa	'pe	'li	du		
<i>final</i>	ẽ	'sã	gɪ	'ko	mwĩa	'pe	'li	du		
<i>final</i>	em	san	gue	co	mo	em	a	pe	li	do.

## Camada 10 (cont.)

**Regra 2524** sinalefa < 0,1‰

**ber:** /⟨ʊiaē⟩/ **msub:** “wjaɟ” **bqsil:** 4 **bpar:** 1

**bton:** (átona, átona, átona, átona) **mpts:** 3

**mdel:** (nada, excluir, excluir, excluir)

MAG.47.7.6 Colocado entre o mundo e a Eternidade,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ko	lo	ka	dwē	trju	mū	du	i	a	e	ter	ni	da	di
<i>depois</i>	ko	lo	ka	dwē	trju	mū	dwjaɟ	ter	ni	da	di			

	1	2	3	4	5	6	7	8	9	10	11					
<i>final</i>	ko	lo	ka	dwē	trju	mū	dwjaɟ	ter	ni	da	di					
<i>final</i>	Co	lo	ca	do	en	tre	o	mun	do	e	a	E	ter	ni	da	de,

**Regra 2533** sinalefa < 0,1‰

**ber:** /⟨εʊ[aiō]⟩/ **msub:** “εX” **bqsil:** 3 **bpar:** 1 **bton:** (tônica, átona, átona)

**mpts:** 5 **mreaj:** (nada, diminuir, nada) **mdel:** (nada, excluir, excluir)

**mmap:** (em 1 inserir 2 com mapa [aiō] → [āiō])

ANJ.17.3.1 É o afrodístico leito do hetairismo,

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'ε	u	a	fro	dis	ti	ku	'lej	tu	dwe	taj	'riz	mu
<i>depois</i>	'εā	fro	dis	ti	ku	'lej	tu	dwe	taj	'riz	mu		

	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	'εā	fro	dis	ti	ku	'lej	tu	dwe	taj	'riz	mu			
<i>final</i>	É	o	a	fro	dís	ti	co	lei	to	do	he	tai	ris	mo,

**Regra 2522** sinalefa < 0,1‰

**ber:** /⟨iε[oā]⟩/ **msub:** “jeX” **bqsil:** 3 **bpar:** 1 **bton:** (átona, tônica, tônica)

**mpts:** 5 **mdel:** (nada, excluir, excluir)

**mmap:** (em 2 inserir 2 com mapa [oā] → [oā])

ANJ.68.1.1 Guerra é esforço, é inquietude, é ânsia, é transporte...

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	'ge	'βεεs	'for	'swēĩ	kje	'tu	dĩ	'ε	ā	'sje	'trās	'por	tĩ
<i>depois</i>	'ge	'βεεs	'for	'swēĩ	kje	'tu	djeā	'sje	'trās	'por	tĩ		

	1	2	3	4	5	6	7	8	9	10	11							
<i>final</i>	'ge	'βεεs	'for	'swēĩ	kje	'tu	djeā	'sje	'trās	'por	tĩ							
<i>final</i>	Gue	rra	é	es	for	ço,	é	in	quie	tu	de,	é	ân	sia,	é	trans	por	te...

## Camada 10 (cont.)

**Regra 2523** sinalefa < 0,1‰

**ber:** /(\uεε)/ **msub:** “wε” **bqsil:** 3 **bpar:** 1 **bton:** (átona, tônica, tônica)  
**mpts:** 5 **mreaj:** (nada, diminuir, nada) **mdel:** (nada, excluir, excluir)

ANJ.279.1.1 D’O Nonevar” no número último, é esta

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	du	no	ne	var	nu	nu	mi	rul	ti	mu	ε	es	ta
<i>depois</i>	du	no	ne	var	nu	nu	mi	rul	ti	mwes	ta		

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	du	no	ne	var	nu	nu	mi	rul	ti	mwes	ta	
<i>final</i>	D’O	No	ne	var	no	nú	me	ro	úl	ti	mo, é es	ta

**Regra 2525** sinalefa < 0,1‰

**ber:** /(\uα[uo])/ **msub:** “wjaX” **bqsil:** 4 **bpar:** 1  
**bton:** (átona, átona, átona, átona) **mpts:** 3  
**mdel:** (nada, excluir, excluir, excluir)  
**mmap:** (em 3 inserir 3 com mapa [uo] → [wɔ])

PIN.83.21.1 “Quem se diz Florentino e à usura é dado,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	kē	si	dis	flo	rē	ti	nu	i	a	u	zu	rē	da	du
<i>depois</i>	kē	si	dis	flo	rē	ti	nwjaw	zu	rē	da	du			

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	kē	si	dis	flo	rē	ti	nwjaw	zu	rē	da	du	
<i>final</i>	“Quem	se	diz	Flo	ren	ti	no	e à u	su	ra é	da	do,

**Regra 2526** sinalefa < 0,1‰

**ber:** /(\uααε)/ **msub:** “wj’ag” **bqsil:** 4 **bpar:** 1  
**bton:** (átona, átona, átona, tônica) **mpts:** 4  
**mdel:** (nada, excluir, excluir, excluir)

ANJ.12.71.2 No homem civilizado, e a ele se prendem

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	nwo	mē	si	vi	li	za	du	i	a	ε	li	si	prē	dē
<i>depois</i>	nwo	mē	si	vi	li	za	dwjaḡ	li	si	prē	dē			

	1	2	3	4	5	6	7	8	9	10	11	
<i>final</i>	nwo	mē	si	vi	li	za	dwjaḡ	li	si	prē	dē	
<i>final</i>	No ho	mem	ci	vi	li	za	do,	e a e	le	se	pren	dem

## Camada 10 (cont.)

**Regra 2527** sinalefa < 0,1‰

**ber:** /ʘje/ **msub:** "wje" **bqsil:** 2 **bpar:** 1 **bton:** (átona, átona) **mpts:** 2  
**mdel:** (nada, excluir)

ANJ.2.3.1 No hierático areópago heterogêneo

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	nu	je	'ra	ti	kwa	'rjo	pa	gwe	te	ro	'ze	nju
<i>depois</i>	nwje	'ra	ti	kwa	'rjo	pa	gwe	te	ro	'ze	nju	

	1	2	3	4	5	6	7	8	9	10	11
<i>final</i>	nwje	'ra	ti	kwa	'rjo	pa	gwe	te	ro	'ze	nju
<i>final</i>	No	hie-	rá-	ti-	co	a-	reó-	pa-	go	he-	te-ro-gê-neo

**Regra 2530** sinalefa < 0,1‰

**ber:** /ʘwae/ **msub:** "jwaɛ" **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, átona)  
**mpts:** 3 **mdel:** (nada, excluir, excluir)

ANJ.139.2.3 — Senhora dai-me u'a esmola — e estertorada

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>antes</i>	se	'jɔ	'ra	'daj	mɪ	wa	ez	'mɔ	'ljes	ter	to	'ra	da
<i>depois</i>	se	'jɔ	'ra	'daj	mjwaɛz	'mɔ	'ljes	ter	to	'ra	da		

	1	2	3	4	5	6	7	8	9	10	11
<i>final</i>	se	'jɔ	'ra	'daj	mjwaɛz	'mɔ	'ljes	ter	to	'ra	da
<i>final</i>	—	Se-	nho-	ra	dai-	me	u'a	es-	mo-	la	— e es-ter-to-ra-da

**Regra 2537** sinalefa < 0,1‰

**ber:** /ʘɪʘe/ **msub:** "je" **bqsil:** 4 **bpar:** 1  
**bton:** (átona, átona, átona, átona) **mpts:** 5  
**mrea:** (nada, diminuir, diminuir, nada)  
**mdel:** (nada, excluir, excluir, excluir)

DUR.10.70.2 Desce do trono e o esplêndido diadema

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	'de	sɪ	'du	'tro	nu	ɪ	u	es	'plẽ	di	'du	dja	'de	ma
<i>depois</i>	'de	sɪ	'du	'tro	njes	'plẽ	di	'du	dja	'de	ma			

	1	2	3	4	5	6	7	8	9	10	11
<i>final</i>	'de	sɪ	'du	'tro	njes	'plẽ	di	'du	dja	'de	ma
<i>final</i>	De-	sce-	do	tro-	no	e	o	es-	plên-	di-	do-dia-de-ma

# Camada 10 (cont.)

**Regra 2545** sinalefa < 0,1‰

**ber:** /(ʊʊ([āa]))/ **msub:** "jw\2" **bqsil:** 4 **bpar:** 1

**bton:** (átona, átona, átona, tônica) **mpts:** 5

**mreaj:** (nada, diminuir, nada, nada) **mdel:** (nada, excluir, excluir, excluir)

ANJ.218.1.2 O calor e o carbono e o amplo éter são

	1	2	3	4	5	6	7	8	9	10	11	12	13			
<i>antes</i>	ʊ	ka	l	r	ju	kar	bo	nu	ɪ	u	ā	plwɛ	tɪr	sāw		
<i>depois</i>	ʊ	ka	l	r	ju	kar	bo	njwā	plwɛ	tɪr	sāw					
<i>final</i>	ʊ	ka	l	r	ju	kar	bo	njwā	plwɛ	tɪr	sāw					
<i>final</i>	O	ca	lo	r	e	o	car	bo	no	e	o	am	plo	é	ter	são

**Regra 2555** sinalefa < 0,1‰

**ber:** /(ejẽ)/ **msub:** "eĩ" **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona) **mpts:** 3

**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

MAT:29.138.2.9 se até agora dei em dar-lhes

	1	2	3	4	5	6	7	8	9
<i>antes</i>	sja	tɛ	a	go	ra	dej	ẽ	dar	ɫɪs
<i>depois</i>	sja	tɛ	a	go	ra	dej	dar	ɫɪs	
<i>final</i>	sja	tɛ	a	go	ra	deĩ	dar	ɫɪs	
<i>final</i>	se	a	té	a	go	ra	dei	em	dar-lhes

**Regra 2511** sinalefa < 0,1‰

**ber:** /(ie)/ **msub:** "iɛ" **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona) **mpts:** 6

**mdel:** (nada, excluir)

MAT:19.4.1.46 temi, emudeci, calei;

	1	2	3	4	5	6	7	8
<i>antes</i>	tɛ	mi	e	mu	dɛ	si	ka	lɛj
<i>depois</i>	tɛ	miɛ	mu	dɛ	si	ka	lɛj	
<i>final</i>	tɛ	miɛ	mu	dɛ	si	ka	lɛj	
<i>final</i>	tɛ	mi	e	mu	dɛ	ci	ca	lei;

## Camada 10 (cont.)

**Regra 2513** sinalefa < 0,1‰

**ber:** /([ãa])/ **msub:** “ãã” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, átona) **mpts:** 6  
**mdel:** (nada, excluir)

CAM.10.19.4 Fará ser vã a braveza com que venha.

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	fa	ra	ser	vã	a	bra	ve	za	kõ	ki	ve	na
<i>depois</i>	fa	ra	ser	vãã	bra	ve	za	kõ	ki	ve	na	
<i>final</i>	fa	ra	ser	vãã	bra	ve	za	kõ	ki	ve	na	
<i>final</i>	Fa	rá	ser	vã a	bra	ve	za	com	que	ve	nha.	

**Regra 2529** sinalefa < 0,1‰

**ber:** /([v]([ãõ]))/ **msub:** “w]2” **bqsil:** 2 **bpar:** 1 **bton:** (átona, tônica)  
**mpts:** 4 **mdel:** (nada, excluir)

ANJ.25.14.2 E enchia com a fluidez do eólico hissopo

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	jẽ	ʃi	a	kwa	fluj	dez	du	jõ	li	kwi	sõ	pɪ
<i>depois</i>	jẽ	ʃi	a	kwa	fluj	dez	dwjõ	li	kwi	sõ	pɪ	
<i>final</i>	jẽ	ʃi	a	kwa	fluj	dez	dwjõ	li	kwi	sõ	pɪ	
<i>final</i>	E	en	chi	a	com	a	flui	dez	do	eó	li	co hi sso pe

**Regra 2531** sinalefa < 0,1‰

**ber:** /([ee])/ **msub:** “ε” **bqsil:** 2 **bpar:** 1 **bton:** (tônica, tônica) **mpts:** 1  
**mreaj:** (nada, diminuir) **mdel:** (nada, excluir)

MAT:6.17.1.7 se você é cobra mandada

	1	2	3	4	5	6	7	8	9
<i>antes</i>	sɪ	vo	se	ε	kõ	bra	mã	da	da
<i>depois</i>	sɪ	vo	se	kõ	bra	mã	da	da	
<i>final</i>	sɪ	vo	se	kõ	bra	mã	da	da	
<i>final</i>	se	vo	cê	é	co	bra	man	da	da



**Camada 10 (cont.)**

**Regra 2532**      sinalefa       $< 0,1\%$

**ber:** /(ɛʁi)/ **msub:** “wi” **bqsil:** 3 **bpar:** 1 **bton:** (tônica, átona, tônica)  
**mpts:** 4 **mreaj:** (nada, nada, diminuir) **mdel:** (nada, excluir, excluir)

ANJ.67.11.3 Esterilizadora de órgãos... É o hino

antes es:te:ri:li:za:do:ra:djɔr gāw zε u ʔi nu  
depois es:te:ri:li:za:do:ra:djɔr gāw zwi nu

final es:te:ri:li:za:do:ra:djɔr gāw zwi nu  
final Es:te:ri:li:za:do:ra:de ór gāo... É o hi: no

**Regra 2535**      sinalefa       $< 0,1\%$

**ber:** /(aɛ)/ **msub:** “aẽ” **bqsil:** 3 **bpar:** 1 **bton:** (tônica, átona, átona)  
**mpts:** 5 **mreaj:** (nada, diminuir, nada) **mdel:** (nada, excluir, excluir)

MAT:4.1.4.2 Acabado já, e em cinzas consumido

antes 1 2 3 4 5 6 7 8 9 10 11 12 13  
a | ka | ba | du | 'za | I | ã | s̃ĩ | zas | kō | su | mi | du

depois a | ka | ba | du | 'zaẽ | s̃ĩ | zas | kō | su | mi | du

final 1 2 3 4 5 6 7 8 9 10 11  
a | ka | ba | du | 'zaẽ | s̃ĩ | zas | kō | su | mi | du

final A- ca- ba- do | já, e em | cin- zas | con- su- mi- do

**Regra 2540**      sinalefa       $< 0,1\%$

**ber:** /(ʁiē)/ **msub:** “jwē” **bqsil:** 4 **bpar:** 1  
**bton:** (átona, átona, átona, átona) **mpts:** 5  
**mreaj:** (nada, nada, nada, disminuir) **mdel:** (nada, excluir, excluir, excluir)

ANJ.272.2.2 Arruine-o, esmaque-o, enforque-o e, em suma, esgane-o

antes 1 2 3 4 5 6 7 8 9 10 11 12 13 14  
a | kuj | njez | ma | gjwē | for | kɪ | u | ɪ | ē | su | mez | gā | nju  
depois a | kuj | njez | ma | gjwē | for | kwē | su | mez | gā | nju

final 1 2 3 4 5 6 7 8 9 10 11  
a | kuj | njez | ma | gjwē | for | kwē | su | mez | gā | nju  
final A-rrui-ne-o, es- ma- gue-o, en- for- que-o e, em- su- ma, es- ga- ne-o

## Camada 10 (cont.)

**Regra 2542** sinalefa < 0,1‰

**ber:** /(ʋɪaɔ)/ **msub:** "jo" **bqsil:** 4 **bpar:** 1  
**bton:** (átona, átona, átona, tônica) **mpts:** 5  
**mreaj:** (nada, nada, diminuir, diminuir)  
**mdel:** (nada, excluir, excluir, excluir)

ANJ.217.1.23 A fumaça, a farelo e a outras fragmentações...

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>antes</i>	a	fu	ma	sa	fa	re	lu	i	a	ow	tras	frag	mẽ	ta	sõjs
<i>depois</i>	a	fu	ma	sa	fa	re	ljow	tras	frag	mẽ	ta	sõjs			

	1	2	3	4	5	6	7	8	9	10	11	12				
<i>final</i>	a	fu	ma	sa	fa	re	ljow	tras	frag	mẽ	ta	sõjs				
<i>final</i>	A	fu	ma	ça	a	fa	re	lo	e	a	ou	tras	frag	men	ta	ções...

**Regra 2544** sinalefa < 0,1‰

**ber:** /(ɪʋɪ([eɛ])/) **msub:** "je" **bqsil:** 4 **bpar:** 1  
**bton:** (átona, átona, átona, tônica) **mpts:** 5  
**mreaj:** (nada, nada, diminuir, diminuir)  
**mdel:** (nada, excluir, excluir, excluir)

ANJ.213.3.4 Abrange-o, e ele é mais triste e ele é mais triste

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	a	brã	ʒɪ	u	i	ɛ	ljɛ	majs	tris	tje	ljɛ	majs	tris	ti
<i>depois</i>	a	brã	ʒje	ljɛ	majs	tris	tje	ljɛ	majs	tris	ti			

	1	2	3	4	5	6	7	8	9	10	11						
<i>final</i>	a	brã	ʒje	ljɛ	majs	tris	tje	ljɛ	majs	tris	ti						
<i>final</i>	A	bran	ge-o,	e	e	le	é	mais	tris	te	e	e	le	é	mais	tris	te

**Regra 2546** sinalefa < 0,1‰

**ber:** /(ɛɔĩ)/ **msub:** "jɛĩ" **bqsil:** 4 **bpar:** 1  
**bton:** (átona, tônica, átona, átona) **mpts:** 5  
**mreaj:** (nada, nada, nada, diminuir) **mdel:** (nada, excluir, excluir, excluir)

ANJ.68.3.1 É a obsessão de ver sangue, é o instinto horrendo

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ɛɔb	se	sãw	dɪ	ver	sã	gɪ	ɛ	u	ɪs	tɪ	two	ɬɛ	du
<i>depois</i>	ɛɔb	se	sãw	dɪ	ver	sã	gɛĩs	tɪ	two	ɬɛ	du			

	1	2	3	4	5	6	7	8	9	10	11						
<i>final</i>	ɛɔb	se	sãw	dɪ	ver	sã	gɛĩs	tɪ	two	ɬɛ	du						
<i>final</i>	É	a	ob	se	ssão	de	ver	san	gue,	é	o	ins	tin	to	ho	rren	do

# Camada 10 (cont.)

**Regra 2547** sinalefa < 0,1‰

**ber:** /(ɪɛai)/ **msub:** “jɛj” **bqsil:** 4 **bpar:** 1

**bton:** (átona, tônica, átona, átona) **mpts:** 5

**mreaj:** (nada, nada, nada, diminuir) **mdel:** (nada, excluir, excluir, excluir)

FIG:5.63.4.3 Ou se é a ilusão do próprio pensamento.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>antes</i>	ow	si	ɛ	a	i	lu	zāw	du	prɔ	prju	pē	sa	mē	tu
<i>depois</i>	ow	sjej	lu	zāw	du	prɔ	prju	pē	sa	mē	tu			
	1	2	3	4	5	6	7	8	9	10	11			
<i>final</i>	ow	sjej	lu	zāw	du	prɔ	prju	pē	sa	mē	tu			
<i>final</i>	Ou	se	é	a	i	lu	são	do	pró	prio	pen	sa	men	to.

**Regra 2549** sinalefa < 0,1‰

**ber:** /(ʋɛaɔ)/ **msub:** “wɔ” **bqsil:** 4 **bpar:** 1

**bton:** (átona, tônica, átona, tônica) **mpts:** 5

**mreaj:** (nada, nada, diminuir, diminuir)

**mdel:** (nada, excluir, excluir, excluir)

ANJ.88.1.1 Cedo à sofreguidão do estômago. É a hora

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
<i>antes</i>	se	dwa	so	fre	gi	dāw	dwes	to	ma	gu	ɛ	a	ɔ	ra	
<i>depois</i>	se	dwa	so	fre	gi	dāw	dwes	to	ma	gwo	ra				
	1	2	3	4	5	6	7	8	9	10	11				
<i>final</i>	se	dwa	so	fre	gi	dāw	dwes	to	ma	gwo	ra				
<i>final</i>	Ce	do	à	so	fre	gui	dão	do	es	tô	ma	go	É	a ho	ra

**Regra 2550** sinalefa < 0,1‰

**ber:** /(iaɪē)/ **msub:** “ijē” **bqsil:** 4 **bpar:** 1

**bton:** (tônica, átona, átona, átona) **mpts:** 3

**mreaj:** (nada, diminuir, diminuir, nada)

**mdel:** (nada, excluir, excluir, nada)

FIG:5.71.3.3 Nessa casa verti-a. E, então, jamais,

	1	2	3	4	5	6	7	8	9	10	11	12
<i>antes</i>	nɛ	sa	ka	za	ver	ti	a	i	ē	tāw	za	majs
<i>depois</i>	nɛ	sa	ka	za	ver	ti	jē	tāw	za	majs		
	1	2	3	4	5	6	7	8	9	10		
<i>final</i>	nɛ	sa	ka	za	ver	ti	jē	tāw	za	majs		
<i>final</i>	Ne	ssa	ca	sa	ver	ti	a	E	en	tão	ja	mais,

## Camada 10 (cont.)

**Regra 2556** sinalefa < 0,1‰

**ber:** /{ɪʋje}/ **msub:** “je” **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, átona)  
**mpts:** 5 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

ANJ.66.1.3 Ante o hieróglifo e o enigma das folhagens,

antes 1 2 3 4 5 6 7 8 9 10 11 12 13  
 'ā tr̩ u je 'rɔ gli fje 'nig ma das fo 'ʎa zēs  
 depois 'ā tje 'rɔ gli fje 'nig ma das fo 'ʎa zēs

final 1 2 3 4 5 6 7 8 9 10 11  
 'ā tje 'rɔ gli fje 'nig ma das fo 'ʎa zēs  
 final An-te o hie-ró-gli-fo e o e-nig-ma das fo-lha-gens,

**Regra 2557** sinalefa < 0,1‰

**ber:** /{ɪaja}/ **msub:** “ja” **bqsil:** 3 **bpar:** 1 **bton:** (átona, átona, átona)  
**mpts:** 6 **mreaj:** (nada, diminuir, diminuir) **mdel:** (nada, excluir, excluir)

ANJ.12.64.2 A estéril terra, e a hialina lâmpada oca,

antes 1 2 3 4 5 6 7 8 9 10 11 12 13  
 aɛs 'tɛ ril 'tɛ ʋɪ a ja 'li na 'lā pa 'dɔ ka  
 depois aɛs 'tɛ ril 'tɛ ʋja 'li na 'lā pa 'dɔ ka

final 1 2 3 4 5 6 7 8 9 10 11  
 aɛs 'tɛ ril 'tɛ ʋja 'li na 'lā pa 'dɔ ka  
 final A es-té-ril te-rra, e a hia-li-na lâm-pa-da o-ca,

**Regra 2558** sinalefa < 0,1‰

**ber:** /{ejaɪa}/ **msub:** “eja” **bqsil:** 4 **bpar:** 1  
**bton:** (tônica, átona, átona, átona) **mpts:** 3  
**mreaj:** (nada, diminuir, diminuir, nada)  
**mdel:** (nada, excluir, excluir, nada)

SIL:1.83.4.1 Eu, triste, olhei-a... e a desventura abranjo

antes 1 2 3 4 5 6 7 8 9 10 11 12 13  
 'ew 'tris tjo 'ʎej a ɪ a dez vē 'tu ra 'brā ʒu  
 depois 'ew 'tris tjo 'ʎej a dez vē 'tu ra 'brā ʒu

final 1 2 3 4 5 6 7 8 9 10 11  
 'ew 'tris tjo 'ʎej a dez vē 'tu ra 'brā ʒu  
 final Eu, tris-te, o-lhei-a... e a des-ven-tu-ra a-bran-jo

### **C. VERSOS MANUALMENTE VERIFICADOS**

Este apêndice lista, a partir da próxima página, todos os 1.080 versos manualmente verificados no experimento principal de validação. Os versos cujo código é precedido de um asterisco foram incorretamente escandidos pelo sistema.

**Corpus CAM: decassílabos**

1.19.5 Da bran- ca es- cu- ma os ma- res se mos- tra- vam  
 1.31.8 De que Ni- sa ce- le- bra in- da a me- mó- ria.  
 1.52.6 Se en- tre vó- s a ver- da- de não se ne- ga,  
 1.53.6 O cla- ro de- scen- den- te de A- bra- ão  
 1.61.8 E mui- to mais con- ten- te co- me e be- be.

1.64.4 De mim, da Lei, da- s ar- mas que tra- zi- a.  
 1.88.1 Qual no co- rro san- gui- no o le- do a man- te,  
 1.90.5 Da ca- val- ga- da ao Mou- ro já lhe pe- sa,  
 1.91.2 Sem for- ça, de co- var- de e de a- pre- ssa- do,  
 1.92.7 Des- ta ar- te o Por- tu- guê- s en- fim cas- ti- ga

1.99.7 A Mo- çam- bi- que es- ta i- lha, que se cha- ma  
 2.39.6 Sem que to me- re- ce- sse, nem te e- rra- sse,  
 2.41.6 Se lhe im- pe- di- ra a fa- la pi- e- do- sa,  
 2.41.7 Tor- na a se- gui- la; e in- do por di- an- te,  
 2.43.1 E com o seu a- per- tan- do o ros- to a- ma- do,

2.49.8 Que, se re- sis- te, con- tra si pe- le- ja.  
 2.63.6 Um Rei, com mui- ta s o- bras de a- mi- za- de,  
 2.91.1 Res- pon- dem- lhe da te- rra jun- ta- men- te,  
 2.102.2 Tu- do o que de seus, Rei- nos lhe cum- pri- sse,  
 2.113.4 Co- nhe- ci- do no mun- do e no- me- a- do:

3.19.2 Su- jei- tan- do Par- té- no- pe in- qui- e- ta;  
 3.25.3 Por- tu- gal h- ou- ve em sor- te, que no mun- do  
 3.65.7 Pe- la fral- da da se- rra, des- cui- da- do  
 3.69.5 Se a té que sem- pre o for- te Rei res- guar- da  
 3.71.4 Ter teu so- gro de ti vi- tó- ria di- na,

3.72.2 E- ní- o- co- s e Col- cos, cu- ja fa- ma  
 3.90.5 No tem- po des- te, aos Mou- ros foi to- ma- do  
 3.105.1 "Por- tan- to, ó Rei, de quem com pu- ro me- do  
 3.122.3 Que tu- do en- fim, tu, pu- ro a- mor, des- pre- za,  
 3.127.5 A es- tas cri- an- ci- nhas tem res- pei- to,

3.130.3 Ma- s o per- ti- naz po- vo, e seu des- ti- no  
 4.20.3 Já pa- ra se en- tre- gar qua- se mo- vi- dos  
 4.38.1 —"Ve- des- me a- qui, Rei vo- sso, e com- pa- nhei- ro,  
 4.76.4 Que a to- dos fo- ram gran- de ad- mi- ra- ção.  
 4.83.1 "Fo- ram de E- ma- nu- el re- mu- ne- ra- dos,

**Corpus CAM: decassílabos (cont.)**

4.92.5	Os	mon-	tes	de	mais	per-	to	res-	pon-	di-	am,
4.95.7	Que	mor-	tes,	que	pe-	ri-	gos,	que	tor-	men-	tas,
5.2.3	E o	mun-	do,	que	com	tem-	po	se	con-	su-	me,
5.10.1	"Po-	ra-	qui	ro-	de-	an-	do a	lar-	ga	par-	te
5.45.2	Com	fa-	ma al-	ta	fi-	zer	to-	ca-	r os	Céus,	
5.45.6	Os	so-	ber-	bo-	s e	prós-	pe-	ros	tro-	féus;	
5.59.4	Por-	es-	tas	lon-	ga-	s á-	guas	se es-	ten-	de-	ram;
5.62.2	Pos-	to	que	to-	do-	s E-	tí-	o-	pe-	s e-	ram,
5.66.6	Que,	co-	mo	tu-	do	ne-	le	são	mu-	dan-	ças.
5.67.8	Com	que	nos	fez	ven-	ce-	r a	grão	co-	rren-	te.
5.69.1	"Des-	ta	gen-	te	re-	fres-	co al-	gum-	to-	ma-	mos,
5.94.8	E-	nei-	a-	s, e a	Ro-	ma-	na	gló-	ria	vo-	e.
6.27.6	O i-	nun-	do u-	ni-	ver-	sal,	e o	tens	cer-	ca-	do,
6.49.6	Com	pa-	la-	vras	de a-	fa-	go-	s e	de a-	mo-	res
6.67.7	En-	quan-	to	se	de-	têm	em	In-	gla-	te-	rra,
6.90.7	De	con-	ten-	te	de	ver	que a	da-	ma o	man-	da,
7.14.1	Ma-	s en-	tan-	to	que	ce-	go-	s o	se-	den-	tos
7.25.5	—"A-	brin-	do,	lhe	res-	pon-	de, o	mar	pro-	fun-	do,
7.29.7	Eu-	rí-	di-	ce,	to-	can-	do a	li-	ra	de ou-	ro,
7.58.4	Que a	seu	cos-	tu-	me es-	ta-	va	ru-	mi-	nan-	do.
7.71.4	Ou	lá	dal-	guns	que	do	Pf-	re-	ne	des-	çam.
7.80.1	A-	go-	ra,	com	po-	bre-	za a-	vo-	rre-	ci-	da,
7.80.4	De	no-	vo,	mais	que	nun-	ca,	de-	rri-	ba-	do;
8.19.2	Con-	tra A-	rro-	n-	ches,	que	to-	ma,	por	vin-	gan-ça
8.49.6	Pois	sa-	be-	rás	que a-	que-	les,	que	che-	ga-	dos
8.74.7	Da	ma-	dre	Té-	tis,	qual	pi-	ra-	ta i-	ni-	co,
8.95.6	Tor-	na-	ra	te-	rra,	por	que	não	pu-	de-	sse
8.96.2	A-	té	ve-	r o	que o	tem-	po	lhe	des-	co-	bre:
8.96.5	Ve-	ja a-	go-	ra o	ju-	í-	zo	cu-	ri-	o-	so
9.32.6	As	cha-	gas	re-	ce-	bi-	das	cu-	ja a-	ju-	da
9.35.4	Tam-	bém	vos	to-	mam	nas	Vul-	câ-	neas	re-	des.
9.40.3	De a-	mor	dos	Lu-	si-	ta-	no-	s in-	cen-	di-	das,
9.48.5	Já	não	fi-	ca	na al-	ja-	va	se-	ta al-	gu-	ma,
9.52.5	Que,	por	que	não	pa-	ssa	ssem,	sem	que	ne-	la
9.54.2	Er-	gui-	dos	com	so-	ber-	ba	gra-	ci-	o-	sa,

**Corpus CAM: decassílabos (cont.)**

	1	2	3	4	5	6	7	8	9	10	11
9.59.2	Cor,	com	que	tu,	ru-	bi,	teu	pre-	ço	per-	des;
9.67.6	De-	fen-	dem	a	ver-	du-	ra,	pa-	sse-	a-	vam
9.78.2	Fu-	gir-	me,	por	que	não	po-	ssa	to-	car-	te,
9.88.3	Nu-	ma al-	ma,	do-	ce, in-	cóg-	ni-	ta a-	le-	gri-	a,
10.7.7	Va-	ti-	ci-	nan-	do, o	di-	sse, e	na	me-	mó-	ria

	1	2	3	4	5	6	7	8	9	10	11
10.11.5	So-	fre-	rá	suas	ci-	da-	de-	s e	lu-	ga-	res,
10.13.3	Rei	de	Co-	chim,	com	pou-	cos	na-	tu-	rais,	
10.16.7	Por	dous	pa-	ssos	num	tem-	po;	mas	vo-	an-	do
10.43.6	Mais	que	li-	ões	fa-	mé-	li-	co-	s e	tou-	ros,
10.58.2	Que	cla-	ra-	men-	te	põem	a-	ber-	to o	ros-	to

	1	2	3	4	5	6	7	8	9	10	11
10.69.2	Tra-	bu-	cos	fe-	ros,	mi-	na-	s en-	co-	ber-	tas,
10.74.6	(Nu-	ma	côn-	so-	na	voz	to-	das	so-	a-	vam),
10.76.5	Se-	gue-	me	fir-	me e	for-	te,	com	pru-	dên-	cia,
10.82.1	"A-	qui,	só	ver-	da-	dei-	ros,	glo-	ri-	o-	sos
10.103.5	Mas	vê	a i-	lha	Ge-	rum,	co-	mo	des-	co-	bre

	1	2	3	4	5	6	7	8	9	10	11
10.140.1	"Mas	cá on-	de	mais	se a-	lar-	ga, a-	li	te-	reis	
10.151.6	Vos	vão	ser-	vir,	com	pa-	sso	di-	li-	gen-	te,

**Corpus MAT: dissílabos**

	1	2	3
9.4.2.9	dois	cor-	nos.

**Corpus MAT: trissílabos**

	1	2	3	4
9.10.49.9	Boa h-	is-	tó-	ria.
9.10.59.9	Boa h-	is-	tó-	ria.
9.10.103.9	Boa h-	is-	tó-	ria.

**Corpus MAT: tetrassílabos**

	1	2	3	4	5
4.35.1.103	que	na	tra-	sei-	ra
36.2.9.4	sen-	do es-	tran-	gei-	ros!
36.2.42.4	a	su-	a	pra-	ta.



**Corpus MAT: pentassílabos**

	1	2	3	4	5	6
9.6.7.1	Clé-	ri-	go	ve-	rão	
9.6.7.8	fur-	ta à	mãe	pu-	to-	na:
9.6.13.5	Cò-	ne-	go ul-	tra-	mar	
9.6.14.3	por	cul-	pas	de a-	trás	
9.6.16.7	e o	mar	lhe	fu-	giu	

	1	2	3	4	5	6
9.6.18.7	nas	ta-	pe-	ça-	ri-	as
10.6.1.9	a-	s a-	ves	ca-	no-	ras
13.1.1.2	os	tem-	pos	se es-	go-	tam,

**Corpus MAT: hexassílabos**

	1	2	3	4	5	6	7
4.4.12.2	Di-	rá	que	te	per-	fi-	lha
5.25.7.6	ve-	nham,	ve-	nham,	ve-	rão	
29.7.5.2.10	Cou-	sas	bem	es-	cu-	sa-	das,

**Corpus MAT: heptassílabos**

	1	2	3	4	5	6	7	8
1.4.2.9	não	são	o-	s e-	fes	que	tem	
1.5.1.51	são, os	que	me a-	cu-	sam	de	da-	nos,
1.9.1.26	es-	crú-	pu-	lo	pe-	que-	ni-	no,
1.14.1.10	po-	rém h-	u-	mil-	de	con-	fi-	o,
1.16.1.5	E	se al-	guém	me	mos-	tra a-	mor	

	1	2	3	4	5	6	7	8
1.16.5.1	Ti-	ra-	ra	cin-	zas	ti-	ra-	nas
1.17.10.4	Bas-	tar-	da,	Ven-	di-	da, In-	jus-	ta.
1.17.12.1	Sa-	zo-	na-	da	ca-	ra-	mu-	nha!
2.1.8.3	e	de-	sse am-	pa-	ro	ren-	di-	dos,
2.1.15.2	de	vo-	sso-	s o-	lhos	com a	gui-	a,

	1	2	3	4	5	6	7	8
2.9.1.8	e	ve-	ne-	no	da	mal-	da-	de?
2.10.3.5	ma-	ta-	ral-	gun-	s a-	pe-	te-	ço,
2.10.4.5	so-	bre	mim	ve-	nham	as	san-	tas
2.11.2.1	Se	não	se-	gues	meu-	s en-	ga-	nos,
2.11.8.6	co-	mo	já	em	ou-	tros	vi,	

**Corpus MAT: heptassílabos (cont.)**

2.17.3.1	Que	des-	cul-	pa	po-	sso	dar,
2.17.5.2	di-	go,	que	dei-	xo o	pe-	ca- do;
2.17.6.3	que	lhe hei	de	mai-	s es-	pe-	rar,
2.21.2.8	pa-	ra	Mãe	su-	a	tão	al- ta,
2.22.3.5	te-	re-	mos	gran-	des	tro-	féus.
2.25.1.10	Deus,	que é	vo-	sso a-	mi-	go	d'al- ma.
2.26.1.3	po-	r i-	sso	se	faz	me-	ni- no,
2.26.3.7	re-	al-	ça	de	Cris-	to o	gos- to:
2.29.5.2	dos	céu-	s a	má-	qui-	na	fe- ra,
2.29.17.10	Es-	cu-	tai	vo-	sso-	s e-	fei- tos
2.29.31.10	Por-	que a	sus-	pi-	ros	de	fo- go.
2.29.53.1	À	me-	sa	do	Sa-	cra-	men- to
2.29.61.4	no	lu-	zi-	men-	to	mais	fi- no:
2.29.65.3	que-	reis	vós	fi-	car	co-	mi- go,
2.29.92.10	Não	de-	pen-	dem	de ex-	pli-	ca- dos.
2.29.96.7	vos	ve-	ne- i	rem	mais	que	tu- do
3.10.3.4	ca-	iu	na-	qui-	lo,	que	fez.
4.6.1.3	por	go-	lo-	tão	de	des-	pa- chos,
4.6.1.64	é	pri-	são	qual-	quer	quar-	ti- nha.
4.8.6.6	seus	do-	ces	ni-	nhos	dei-	xa- ram,
4.9.1.44	to-	ca a-	qui,	to-	ca a-	li	sig- nos.
4.10.1.30	no	de-	cur-	so	de	trê-	s a- nos
4.16.3.5	de es-	tre-	la	tão	im-	por-	tu- na
4.25.1.1	Vei-	o ao:	Es-	pí-	ri-	to	San- to
4.25.3.1	Pa-	riu	a	seu	tem-	po um	cu- co,
4.25.6.3	e e-	ra	cri-	a-	da	das	Da- mas
4.25.8.6	pe-	los	que en-	for-	car	fa-	zi- a,
4.25.9.7	e	se	mai-	or	ma-	le-	fi- cio
4.25.10.3	mas	que	me	man-	de en-	for-	car
4.32.1.42	quan-	do	de-	sen-	te-	rram	mor- tos
5.2.2.9	sem-	pre o	pão:	quo-	ti-	di-	a- no,
5.2.4.1	Há	cou-	sa	co-	mo es-	tar	ven- do
5.2.5.7	que	Deus	(co-	mo en-	si-	na a	fé)
5.2.6.7	que	na	Or-	dem	ma-	s aus-	te- ra
5.9.5.8	que u-	ma	fa-	ís-	ca	ca-	iu,

**Corpus MAT: heptassílabos (cont.)**

5.10.3.8 da mú- si- ca os se- te sig- nos,  
 5.12.4.7 e a sem- ra- zão, que se a- sso- ma,  
 5.14.1.5 ca- da qual po- de in- fe- rir,  
 5.18.5.6 diz: fa- ço, por- que fa- çais,  
 5.21.2.10 pe- la qual Deu- s o cas- ti- ga:

5.21.3.3 so- bre qua- tro, que já traz,  
 5.21.3.10 vai es- ta pon- ta à ba- rri- ga:  
 5.23.6.9 que a bo- ti- ca te da- ri- a  
 5.24.1.46 a- du- la- dor sem re- bu- ço,  
 5.24.1.100 on- de es- tá Pá- ro- co in- tru- so:

5.31.3.6 que com tan- ta ma- ti- na- da  
 5.38.3.1 Se lou- vo vo- ssa al- ta i- déi- a,  
 5.39.1.5 o so- ber- bo ao sol a- ris- co  
 5.39.1.6 se põe, e o hu- mi- lha- do não,  
 5.40.1.10 vo- mi- tas- te- s u- ma as- ni- a:

5.42.5.10 a be- rrar pe- las ca- si- nhas.  
 5.42.12.5 na câ- ma- ra es- tás pro- vi- do  
 5.43.1.8 que ren- do a ca- bra le- var,  
 5.44.3.5 vós lo- go me des- men- tis- tes,  
 5.44.11.8 co- mo a- qui sem- pre fi- zes- tes,

5.46.5.10 ten- des cer- to o cen- te- ná- rio.  
 5.47.4.7 é, por- que ce- go do ví- cio  
 \*6.3.1.76 de um par- ce mih- i can- ta- do.  
 6.4.4.4 vos so- be- ja o co- ra- ção:  
 6.7.1.7 por- que en- tão a não fa- ri- a,

6.8.3.5 diz, que pa- ra Por- tu- gal  
 6.11.1.7 poi- s o mai- o- r ao me- nor  
 6.12.1.9 que pa- ra vós, foi tra- gé- dia,  
 6.12.6.6 e tem gran- dí- ssi- mo me- do,  
 6.13.5.3 on- de o jo- go vos de- rri- ba

6.15.7.1 A los Mo- ros por di- ne- ro  
 6.15.7.9 que aos to- lei- rões por di- nhei- ro,  
 6.18.2.10 quan- do lhe ti- rou a pro- va.  
 6.19.5.4 por- que sois bom te- ne- dor:  
 6.19.6.1 Po- r es- ta te- rra se con- ta,

**Corpus MAT: heptassílabos (cont.)**

	1	2	3	4	5	6	7	8	
7.9.1.5	de an-	ti-	no-	má-	sia o	Se-	nhor,		
7.9.3.5	o	rai-	o	do	Sol,	que es-	cal-	da,	
7.11.4.1	Tan-	to a-	li	se en-	trís-	te-	ce-	ram,	
7.11.6.7	e os	que i-	ma-	gi-	nam,	que é	ba-	rro	
7.12.2.6	pal-	ma-	das	tão	bem	dis-	pos-	tas,	
	1	2	3	4	5	6	7	8	
8.3.2.7	com	i-	gual	mão,	i-	gual	ar-	te	
9.2.6.1	Des-	te-	s a-	va-	ros	mo-	fi-	nos,	
*9.2.6.3	de	to-	da a	i-	gua-	ria i-	sen-	ta,	
9.3.3.10	eu	re-	nun-	ci-	o à	for-	tu-	na.	
9.4.2.4	na	ba-	rri-	ga a	Re-	la-	ção:		
	1	2	3	4	5	6	7	8	9
9.4.3.4	dá	fé,	e	co-	bra o	sa-	lá-	rio:	
9.5.1.39	hei	de	fa-	lar	por	pri-	mei-	ro,	
9.5.1.173	E	di-	zen-	do o	ju-	be	dó-	mi-	ne
9.10.2.4	mais	que	na-	sce-	r, e	mo-	rrer:		
9.10.20.7	não	dêem	con-	ta ao	con-	fe-	ssor		
	1	2	3	4	5	6	7	8	
9.10.26.4	so-	bre	dá	cá a-	que-	la	pa-	lha:	
9.10.26.5	que	quei-	ra em	ci-	vil	ba-	ta-	lha	
9.10.27.2	pa-	ra	co-	me-	te-	rem	ví-	cios,	
9.10.47.4	des-	tru-	in-	do o	ca-	be-	dal:		
9.10.61.5	que	sen-	do	pou-	co	fi-	éis		
	1	2	3	4	5	6	7	8	
9.10.63.3	do	be-	ne-	fi-	cio, ou	pen-	são,		
9.10.87.4	e	não	co	de-	mo a-	mi-	ga-	do:	
9.10.88.5	que es-	te	mes-	mo	com	cor-	co-	va	
9.10.91.2	an-	dem	fei-	tos	tran-	ca-	ru-	as	
9.10.97.7	aos	tais	sir-	va	de	tor-	men-	to	
	1	2	3	4	5	6	7	8	
10.1.2.1	On-	de eu	vi	du-	a-	s Au-	ro-	ras	
10.1.3.6	que	dão	mais	for-	ça à	ques-	tão,		
10.21.6.8	que on-	de	quer,	que	pu-	nha o-	s o-	lhos,	
10.32.1.31	pa-	ra	vi-	ver,	quem	te a-	do-	ra,	
10.32.1.46	os	di-	as,	ma-	nhã	s, e	tar-	des,	
	1	2	3	4	5	6	7	8	
11.1.1.7	mais	be-	li-	co-	sa,	que o	fo-	go,	
12.3.5.7	que	di-	sse,	que o	seu	tar-	dar		
12.4.1.6	e	de	vi-	nho a	mes-	ma	có-	pia,	
12.5.2.6	os	ma-	tos	mai-	s es-	con-	di-	dos	
12.5.17.4	foi	ao	a-	tar	das	fe-	ri-	das:	

**Corpus MAT: heptassílabos (cont.)**

12.6.7.10	mais	sus-	to,	que u-	ma es-	pin-	gar-	da.
12.6.9.4	e a	pra-	ça	vei-	ga	de a-	rei-	a:
12.6.9.6	de i-	ni-	mi-	ga	tão	co-	mu-	a,
12.9.3.5	mas	po-	r i-	sso a-	ven-	te-	ja-	do
12.10.3.2	mu-	lhe-	res	de	to-	do o es-	ta-	do,
12.10.7.1	l-	ná-	cio é	gran-	de es-	tu-	dan-	te,
12.12.1.9	de-	pois	de	tan-	ta	ra-	zão	
12.14.2.8	fa-	zi-	am	com	gra-	ças	su-	mas
12.14.18.4	se em-	pras-	ta-	do o	mes-	mo	val:	
13.4.1.49	Des-	de-	nhar-	se a	meus	ca-	ri-	nhos,
13.8.1.6	por-	que	ne-	scia-	men-	te	tro-	co
13.8.1.33	Ser	con-	fi-	a-	do	me a-	len-	ta,
13.8.1.93	Mar-	fi-	da, eu	mo-	rro, eu	a-	ca-	bo:
13.10.3.8	a	fé,	com	que	vo-	s a-	do-	ro,
13.10.7.6	tão	su-	bi-	do é	o a-	rre-	bol,	
13.20.4.7	em	mãos	tão	su-	pe-	ri-	o-	res,
13.25.1.7	a	ca-	da ins-	tan-	te	sus-	pi-	ro,
15.1.2.8	é	cos-	tu-	me	da-	s l-	gre-	jas.
15.10.6.9	por-	que es-	ta	mer-	da	foi	mer-	da,
15.10.9.8	tan-	to	fe-	deis,	que ao	na-	riz	
15.11.4.5	que	pie-	do-	so-	s, e	be-	nig-	inos
15.13.7.5	ho-	je um	re-	ce-	be	so-	men-	te:
15.16.3.9	co-	men-	do	das	vo-	ssas	par-	tes,
15.18.1.7	que	me	ga-	nhai-	s es-	tan-	dar-	tes:
15.18.5.3	a	ser	mem-	bro	da	jus-	ti-	ça,
16.1.1.21	A	mi-	nha	fo-	ra a	pri-	mei-	ra,
16.3.1.54	nem	que o	man-	de a	mi-	nha a	mi-	ga,
16.5.2.8	por-	que	sei,	que	na	Bah-	i-	a
16.5.3.6	que eu	co-	rro, é	se-	ra-	çou-	ta-	do
16.7.1.12	se é	tão	for-	ço-	so, e	tão	be-	lo.
16.9.1.2	ne-	nhum	sus-	to	me	tem	da-	do,
16.11.3.9	e	já	que	fos-	te	bran-	dú-	zia,
17.2.1.46	faz	mu-	dan-	ça a	na-	tu-	re-	za,
17.3.1.2	An-	to-	ni-	ca à	vo-	ssa	ca-	sa,
17.3.1.26	on-	de um	fa-	vor,	que	se	gas-	ta,

**Corpus MAT: heptassílabos (cont.)**

18.1.17.1	1	2	3	4	5	6	7	8
	Por	de-	bai-	xo	de u-	ma a-	ma-	rra
18.2.10.7	chei-	rar-	lhe a-	rou-	pa a-	mon-	don-	go
18.3.4.4	te-	men-	do,	que	te	não	ga-	bem:
19.3.1.26	quan-	do e-	la	do-	s o-	lhos	pu-	xa,
19.4.1.41	Eu	quis	des-	co-	brir-	lhe o a-	mor	
19.4.1.57	1	2	3	4	5	6	7	8
	Eu	de-	ssa e-	lei-	ção	a-	pe-	lo,
19.6.1.1	Que	to-	do o	bem	se	fa-	ri-	a
19.7.5.7	hei	de	des-	pre-	za-	r o a-	gou-	ro,
22.10.1.10	por-	que an-	dam	co	pa-	po	fei-	to?
22.15.1.56	é	ân-	co-	ra	de	que-	rer-	vos.
22.15.1.69	1	2	3	4	5	6	7	8
	Não h-	a-	ve-	rá	tem-	pes-	ta-	de,
22.21.1.9	com	o	san-	gue a-	la-	ga-	di-	ço
23.2.2.8	de	to-	do o	po-	vo	mal-	va-	do
23.2.6.5	ma-	s a	gen-	te a-	li	che-	ga-	da
23.2.7.8	ór-	fão	tão	de-	sam-	pa-	ra-	do,
23.2.15.5	1	2	3	4	5	6	7	8
	mas	so-	bre-	vin-	do um	fe-	brão	
23.2.19.4	cho-	ra-	va	co-	mo um	ju-	men-	to:
23.4.4.1	Po-	rém	de-	pon-	do a-	rro-	gân-	cia
23.5.1.56	vos	tre-	san-	da o	cham-	ba-	ril.	
24.4.3.1	ls-	to a-	gra-	ve-	za	ca-	sou	
24.6.11.4	1	2	3	4	5	6	7	8
	por-	que	vós	sois	Mo-	ça	be-	la,
24.13.1.4	com	tão	ne-	gro-	s a-	ze-	vi-	ches.
24.14.2.3	que	não	po-	de	Mal-	me-	que-	res
24.14.4.2	que a-	ba-	lar	po-	ssa es-	ta	fé,	
24.19.6.3	quan-	do	vó-	s em	pon-	tos	d'hon-	ra
24.20.1.2	1	2	3	4	5	6	7	8
	to-	do o	meu	co-	ra-	ção	sen-	te,
25.2.3.9	com	ba-	rre-	te	de	fa-	lu-	a,
25.3.1.33	O	que eu	qui-	se-	ra,	Be-	ti-	ca,
25.5.4.6	que	per-	deu	na	su-	ja	gue-	rra;
25.6.3.1	Que há	de	te-	r em	gran-	de a-	pre-	ço
25.10.3.10	1	2	3	4	5	6	7	8
	com	pu-	ta	que	te	pa-	riu.	
26.6.1.5	quan-	to	me-	lho-	r ad-	ver-	ti-	da
26.6.3.3	e o	que é	ga-	la	na	ma-	nha,	
26.11.1.5	vi-	vo-	s, e	lo-	go	vos	quis	
26.24.10.5	a-	ssim	crei-	o, a-	ssim	con-	fi-	o,

**Corpus MAT: heptassílabos (cont.)**

26.24.10.6	1	2	3	4	5	6	7	8
	ne-	le	me hei	de	con-	ver-	ter,	
26.25.2.5	pe-	na,	que	mais	se	du-	pli-	ca,
26.25.3.10	Qual	te-	rá	mai-	or	pe-	sar.	
26.26.2.10	Qual	e-	ra a	mor-	te	ti-	ra-	na.
26.30.1.10	Não	que-	ro, o	que	vós	que-	reis.	
26.31.1.1	1	2	3	4	5	6	7	8
	Vi-	di-	nha:	por	que	cho-	rais?	
27.1.1.59	ten-	do a	ca-	ra-	ça	re-	don-	da
27.1.1.116	por-	que	bem	sa-	be em	qual-	quer.	
27.10.2.4	ou	to-	má-	la	na	gai-	o-	la:
28.3.1.8	lo-	go	de	ver-	de	se en-	lu-	ta,
29.1.1.89	1	2	3	4	5	6	7	8
	Vi-	am-	se a-	s en-	fer-	ma-	ri-	as
29.5.3.2	e a	vis-	ta	não	di-	ver-	ti-	a,
29.10.4.3	te-	nha	po-	r es-	pe-	lho a o-	gos-	to
29.12.4.4	que	se-	jais	meu	bran-	co or-	de-	na:
29.15.6.3	por-	que um	a-	mor,	que é	tão	for-	te,
29.22.7.6	1	2	3	4	5	6	7	8
	que es-	tan-	do a	to-	da-	s ex-	pos-	ta,
29.28.1.18	no	mun-	do ou-	tro	par-	não	têm,	
29.30.4.5	vo-	s en-	for-	can-	do	no	ar	
29.30.8.4	por	fo-	ra	tu-	do	per-	deis:	
29.30.8.6	ao	de-	mo, a	quem	vo-	s eu	dou,	
29.35.2.5	1	2	3	4	5	6	7	8
	a-	ssim	me in-	dul-	gen-	cia-	reis	
29.35.3.1	Mal-	di-	ta	se-	ja a	ca-	ra-	ça,
29.36.1.10	ou	se	de	Pai	o e-	xce-	deis.	
29.36.2.9	que en-	tre	tan-	to	mal	cu-	ra-	do,
29.36.3.10	foi	con-	tá-	gio	de	mil	sãos.	
29.36.4.8	1	2	3	4	5	6	7	8
	com	pe-	ri-	go	tão	a-	troz	
29.38.1.8	por-	que é	gran-	de es-	pa-	da-	chim.	
29.40.3.10	o-	brai-	s en-	dia-	bra-	da-	men-	te.
29.58.9.9	sen-	ti-	do	de eu	lhos	plan-	tar,	
29.59.1.1	Se-	nho-	r: o	vo-	sso	ta-	ba-	co
29.61.1.3	1	2	3	4	5	6	7	8
	que a um	Fer-	nan-	des	não	é	da-	do,
29.62.4.1	Na-	da	dis-	to	lhe	va-	leu,	
29.63.9.7	e	tam-	bém	po-	sso a-	fir-	mar,	
29.71.9.7	por	cer-	ta	vi-	a	se-	cre-	ta;
29.71.13.1	Que	vis-	ta,	quem	ren-	das	tem,	

**Corpus MAT: heptassílabos (cont.)**

29.71.15.6	em	pa-	lá-	cio	sem	bo-	le-	ta,
29.76.1.71	es-	te,	que	quei-	mou	a	Trói-	a,
29.76.1.163	um	em-	bas-	ba-	car	na	vis-	ta,
29.77.10.9	quem	so-	men-	te	me-	re-	ci-	a,
29.77.11.9	guar-	de-	se a	cruz	do	di-	a-	bo,
29.82.3.9	se	com a	pre-	ta é	des-	gra-	ça-	do
*29.90.3.9	co-	ma eu	da	frui-	ta	en-	fim,	
29.93.7.10	a	jus-	ti-	ça,	jus-	ti-	ça-	da.
29.98.1.1	Pe-	los	nai-	pes	da	ba-	ra-	lha
29.98.1.24	e	diz,	se	jo-	ga, eu,	me	ga-	nho.
29.98.1.27	que	vão	à-	s al-	mas	fa-	zen-	do,
29.104.1.11	Pe-	go	na es-	co-	ta	com a	mão,	
29.108.1.48	co-	mo a	par-	te é	tão	rei-	mo-	sa,
29.108.1.55	com	me-	cha	lho-	s en-	cai-	xei,	
29.108.1.62	po-	rém	es-	ta,	que	se a-	fas-	ta,
29.110.1.10	fol-	gou	com a	pa-	ter-	ni-	da-	de.
29.111.3.1	Nes-	ta he-	re-	si-	a	tão	cra-	ssa
29.118.1.43	co-	mo	se	viu	no	seu	cen-	tro,
29.118.1.57	que-	r an-	tes,	co-	mo	se	vê,	
29.122.1.35	e	va-	mo-	s ao	ma-	ta-	lo-	te
29.128.1.10	um	lu-	gar	de	bus-	ca-	pés.	
29.128.1.33	O	que e-	ram	al-	tas	mu-	ra-	lhas,
29.128.1.69	nin-	guém	vo-	s a-	cei-	ta o	va-	so,
29.129.1.20	não	pe-	ças	mais,	meu	An-	ji-	nho.
29.132.1.10	e-	ra	for-	ça,	se a-	cen-	de-	sse.
29.132.1.13	e e-	le	quis	vo-	s es-	fre-	gar	
29.132.1.21	Vó-	s es-	ten-	des-	te-	s a	mão,	
29.132.1.39	a-	trás	do	co-	mer	co-	çar,	
29.135.1.68	e	bem	que as	de-	ram,	con-	tu-	do
29.136.2.5	a	mim	não	me	ma-	ra-	vi-	lha,
29.137.2.46	tão	mes-	tra	da	for-	mo-	su-	ra,
29.138.3.2	não h-	ou-	ve	fal-	ta, an-	tes	so-	bre,
29.139.1.32	nem	crei-	a	de	mim,	que	sin-	ta,
30.6.2.10	To-	do o	meu	me-	ti	em	ti	
30.6.3.5	e	quan-	do a	gue-	rra	se a-	tei-	a,



**Corpus MAT: heptassílabos (cont.)**

30.6.4.7	e	sen-	do	fron-	tei-	ro a-	li
30.8.1.5	vós	sem-	pre a	me a-	bo-	rre-	cer
30.11.1.8	co-	mo	me	po-	de es-	que-	cer?
30.11.1.67	co-	mo	vós	sois	tes-	te-	mu-nha,
30.13.1.43	por-	que o	jim-	bo,	que	pe-	diu,
30.16.1.95	en-	tre-	tan-	to	Deus	te	guar-de,
30.17.1.10	On-	tem	sou-	be o	vo-	sso	mal.
31.1.2.10	Tal	fi-	quei,	que	des-	de en-	tão.
31.3.7.5	que es-	pe-	ra-	va a	Ta-	ba-	ro-a
32.1.1.8	quei-	xei-	me,	vol-	tou-	se em	pe-nha.
33.2.5.5	por-	que em	sen-	do	con-jun-	tu-	ra,
33.7.4.5	tan-	ta	pi-	men-	ta	res-	cal-do,
33.9.3.5	sen-	tiu-	se	Jo-	a-	na	tan-to
33.15.1.5	fos-	tes	tão	gran-	de	ma-	ga-na
33.16.3.6	mas	dai	ao-	s o-	lhos,	que	ver,
33.16.6.7	es-	te es-	ti-	lo é	já	tão	ve-lho
33.19.2.9	é	por-	que	na	rea-	li-	da-de
33.25.2.9	que hon-	ras	de	Per-	na-	me-	rim
33.28.1.33	A-	ssim	o	fa-	çam	meus	Fi-lhos,
33.31.1.83	a	ca-	sa	de	vo-	sso a-	mi-go,
33.34.7.10	não	vos	le-	ve a-	s o-	bras	mor-tas.
33.35.3.8	por-	que	sou	um h-	o-	mem	tal,
33.38.4.9	en-	tão	mui	su-	jo o	tra-	gais
33.39.3.6	são	fal-	sá-	rios,	des-	le-	ais;
33.41.4.9	eu	nem	da	mor-	te	sou	dig-no,
34.1.1.3	ou	l-	ná-	cia, ou	A-	po-	lô-nia,
34.2.1.8	to-	do o	dia,	em	que	lhe	pe-se.
34.2.1.39	que	de	mui	se-	ca	re-	cei-o,
34.4.1.4	a	bo-	qui-	nha	de	Po-	lô-nia.
35.2.3.3	poi-	s em	con-	ti-	nos	ge-	mi-dos
35.5.1.1	Que	fe-	bre	têm	tão	ti-	ra-na
36.4.1.95	a-	ve-	sso	de	to-	do o	mun-do,
37.1.4.4	quan-	do es-	ti-	ver	mais	ca-	la-do:
37.3.1.45	Hou-	ve,	que a-	pós	da	sen-	ten-ça,
37.4.18.1	Lo-	bos	ser-	vis,	car-	ni-	cei-ros,

**Corpus MAT: heptassílabos (cont.)**

	1	2	3	4	5	6	7	8
37.4.41.3	co-	mo	fe-	bre-	s aos	mo-	lha-	dos
38.21.1.42	nem	se	de Hu-	ma-	na	tem	no-	me;
38.23.3.10	sem-	pre os	da-	dos	me	per-	de-	ram.
39.5.1.3	as	mais	bre-	ves	trê-	s au-	ro-	ras,
39.6.5.4	que a-	té a	mor-	te	guar-	da-	rá:	

	1	2	3	4	5	6	7	8
39.7.1.9	"Quem	pa-	issa,	Lou-	ro,	quem	pa-	

**Corpus MAT: eneassílabos**

	1	2	3	4	5	6	7	8	9	10
10.6.3.1	Poi-	s o;	Céu,	os,	Pla-	ne-	ta-	s, e	Es-	tre-
										ilas

**Corpus MAT: decassílabos**

	1	2	3	4	5	6	7	8	9	10	11
2.3.3.1	O	pei-	to	já	foi	par-	te en-	tre as;	do;	to-	do,
2.8.1.3	Por-	que	quan-	to	mais	te-	nho	de-	lin-	qüi-	do,
2.18.1.1	O	a-	le-	gre	do	día	en-	tris-	te-	ci-	do,
2.24.2.4	Des-	tes	Fi-	lhos	Ter-	cei-	ros,	que	cri-	as-	te.
3.6.2.3	Ou	to-	da,	por-	que in-	ve-	ja	não	ti-	ve-	ra

	1	2	3	4	5	6	7	8	9	10	11
4.2.2.1	O	céu	vei-	o a	lo-	grá-	la,	mas	por	sor-	te,
4.4.6.5	Can-	ga-	lhas,	que	for-	ma-	ram	lu-	mi-	no-	sas
4.4.7.6	Que em	te o-	lhan-	do,	não	ve-	jo	mais	que um	bu-	rro.
4.4.8.5	A-	té	que	se es-	con-	de, on-	de	mal	o	ve-	jo
4.12.2.2	Aos	ri-	go-	res	do	tem-	po	pós	ren-	di-	do

	1	2	3	4	5	6	7	8	9	10	11
4.27.3.2	Ne-	fan-	do	gos-	to	tem	o	seu	ca-	rá=,	
4.29.2.4	A	vi-	a	bus-	que,	que o	ne-	gó-	cio é	ca-	no.
4.36.1.4	Pás-	coa	fi-	cou	de	re-	den-	ção	o	di-	a.
4.36.2.4	Er-	go	sois	vós,	Se-	nhor,	Deus	da	Bah-	i-	a.
7.1.30.1	A-	ssim	di-	sse, e	pa-	rou,	e eu	a-	ssim	fa-	ço,

	1	2	3	4	5	6	7	8	9	10	11
9.7.4.2	Que, a	quem	de	pro-	fi-	ssão	fal-	ta à	ver-	da-	de,
10.8.2.4	Que	fi-	ar-	me	de um	né-	scio a-	tre-	vi-	men-	to.
13.2.2.1	En-	quan-	to	com	gen-	til	des-	cor-	te-	si-	a
13.12.3.3	Pa-	gan-	do-	me an-	te-	mão,	quan-	to	me-	re-	ço.
16.2.3.1	De	que	po-	de	ser-	vir,	se-	res	mais	be-	la,

**Corpus MAT: decassílabos (cont.)**

21.2.2.3	Man-	de,	que	lho	pa-	ssei-	e a	mor	cau-	te-	la,
22.5.4.1	Pois	to-	da a	Mãe	seus	Fi-	lho-	s a-	co-	mo-	da,
22.6.4.2	Co-	mo	de	si as-	que-	ci-	da a	mão	se a-	tre-	ve
24.3.6.7	E-	fei-	to,	que é	mal	fei-	to ao	di-	zer	mais,	
26.5.1.1	Di-	to-	so	tu,	que	na	pa-	lho-	ça a-	gres-	te
26.23.3.3	E	tu-	do es-	tá	su-	jei-	to a	mil	mu-	dan-	ças.
29.48.3.4	que en-	tre o-	s ho-	mens	não	a-	chas	mais	que es-	có-	ria:
30.14.7.1	Bem	se	zom-	ba	do	po-	bre	fo-	ra-	gi-	do,
33.1.1.1	Há	cou-	sa	co-	mo es-	ta-	r em	São	Fran-	cis-	co,
33.11.3.5	Ó in-	faus-	tas	mo-	ças	na	mo-	fi-	na	ra-	ras,
33.11.9.4	En-	fei-	ta a	ca-	ra	bem,	pen-	teia	o	pê-	lo,
33.11.12.6	A-	mor,	mi-	nha	se-	nho-	ra, é	to-	do e-	xce-	sso.
33.11.14.6	Fi-	ca	so-	bre um	pe-	ne-	do, ou-	tro	pe-	ne-	do.
38.13.2.4	Que	se-	ja o	bri-	o ob-	je-	to do	cui-	da-	do.	
38.16.2.1	Com	ra-	zão	o	res-	pei-	to	só	me a-	gra-	da,
38.24.1.4	Que	mo-	rrí-	eis	por	ver-	vo-	s e-	ixclu-	f-	da.
38.24.2.3	Fin-	gir,	que	po-	r a-	mo-	r, ou	por	pie-	da-	de

**Corpus MAT: hendecassílabos**

15.12.3.2	Que e-	sse	lhe	dá	de	ren-	de-	r o	pão	da	mê-,
-----------	--------	-----	-----	----	----	------	-----	-----	-----	----	------

**Corpus DUR: decassílabos**

1.5.8	Vi-	rei-	s a	ser	mo-	nar-	ca	de ou-	tro	mun-	do.
1.8.8	O	vo-	sso	tro-	no em	pro-	pa-	gar-	se a l-	gre-	ja.
1.54.1	En-	tão	o	bom	mi-	nis-	tro: "É	jus-	to, a-	mi-	go,
1.77.1	Tal	do	ca-	bo-	clo	foi	a	fú-	ria in-	fan-	da;
1.85.2	Com	voz	so-	no-	ra a	com-	pa-	nhia	a-	ni-	ma,
2.4.7	E	tur-	ba	vil	fa-	zen-	do em	mil	pe-	da-	ços,
2.85.6	E-	la,	se	não	me en-	ga-	na o	que al-	ma es-	pe-	ra,
*3.4.5	"—	Diz	me, hós-	pe-	de	a-	mi-	go,	se	pro-	fe-
3.14.2	Um	pen-	sa-	men-	to	d'al-	ma eu	di-	vi-	di-	ra!
3.34.2	Que	com	a	ra-	ma e	flor	for-	mam	pa-	ssei-	os,

**Corpus DUR: decassílabos (cont.)**

3.39.1	1	2	3	4	5	6	7	8	9	10	11
	A-	li	dos	pai-	s a-	s al-	mas	ven-	tu-	ro-	sas
3.39.8	1	2	3	4	5	6	7	8	9	10	11
	Pa-	ssam	tran-	qüi-	la a	noi-	te e a-	le-	gre o-	di-	a.
3.44.7	1	2	3	4	5	6	7	8	9	10	11
	Em	bre-	ve há-	s de	sen-	ti-	r, e	sem	re-	mé-	dio,
3.63.5	1	2	3	4	5	6	7	8	9	10	11
	Ra-	ro	quem	o a-	do-	re ou	quem	o a-	me;		
3.65.3	1	2	3	4	5	6	7	8	9	10	11
	Pon-	do	su-	plí-	cios	con-	tra os	mau-	s ex-	tre-	mos,
3.81.7	1	2	3	4	5	6	7	8	9	10	11
	E	na	san-	ta	dou-	tri-	na,	que en-	si-	na-	va,
4.3.5	1	2	3	4	5	6	7	8	9	10	11
	A-	li	pa-	ssa-	r a ar-	den-	te	ses-	ta or-	de-	na,
4.8.6	1	2	3	4	5	6	7	8	9	10	11
	Ao	la-	do	de	seus	pai-	s, a	be-	la	fi-	lha.
4.53.7	1	2	3	4	5	6	7	8	9	10	11
	Que	tan-	to es-	tra-	go	faz,	que	qual-	quer	fe-	ra
4.68.7	1	2	3	4	5	6	7	8	9	10	11
	On-	de en-	tran-	do o i-	ni-	mi-	go,	sem	pre-	vê-	lo,
5.6.1	1	2	3	4	5	6	7	8	9	10	11
	Quan-	tos	cri-	ar	pu-	de-	ra	que o	ser-	vi-	ssem,
5.32.8	1	2	3	4	5	6	7	8	9	10	11
	O	grão-	Ca-	ra-	mu-	ru,	que o	fo-	go en-	can-	ta.
5.45.6	1	2	3	4	5	6	7	8	9	10	11
	Um	di-	lú-	vio	de fo-	go	n'á-	gua ar-	di-	a,	
6.11.7	1	2	3	4	5	6	7	8	9	10	11
	Quan-	tos	pro-	dí-	gio-	s o	lu-	ga-	r en-	ce-	rra,
6.13.2	1	2	3	4	5	6	7	8	9	10	11
	Pe-	lo es-	tra-	nho es-	pe-	tá-	cu-	lo	sus-	pen-	so,
6.16.3	1	2	3	4	5	6	7	8	9	10	11
	Pois	de	to-	do	sal-	var	ten-	des	von-	ta-	de,
6.21.4	1	2	3	4	5	6	7	8	9	10	11
	Dis-	pa-	ra ao	tem-	po	mes-	mo a ho-	rrí-	vel	bom-	ba.
6.27.2	1	2	3	4	5	6	7	8	9	10	11
	Vi-	mos,	de-	scen-	do	ri-	o, em	cur-	so i-	men-	so,
6.28.3	1	2	3	4	5	6	7	8	9	10	11
	Per-	se-	gui-	dos	de	tro-	pas	nu-	me-	ro-	sas,
6.45.3	1	2	3	4	5	6	7	8	9	10	11
	Os	ca-	sos	de	Di-	o-	go,	que es-	cu-	ta-	va,
6.63.4	1	2	3	4	5	6	7	8	9	10	11
	A en-	vi-	a-	r um	cos-	mó-	gra-	fo,	que ex-	per-	to
6.64.6	1	2	3	4	5	6	7	8	9	10	11
	Quan-	do	com	pei-	to	se a-	tra-	ve-	ssem	for-	te
6.70.1	1	2	3	4	5	6	7	8	9	10	11
	Eu	de-	pois	des-	tes	na o-	ca-	sião	pre-	sen-	te,
7.22.6	1	2	3	4	5	6	7	8	9	10	11
	E-	xa-	ta	na-	rra	ção	de	có-	pia in-	gen-	te,
7.47.4	1	2	3	4	5	6	7	8	9	10	11
	E	ren-	de	no-	vo	fru-	to em	ca-	da	lu-	a;
7.72.3	1	2	3	4	5	6	7	8	9	10	11
	Com	a-	sa e	cau-	da os	lí-	qui-	do-	s es-	pa-	ços
7.73.6	1	2	3	4	5	6	7	8	9	10	11
	Com	que a-	tra-	in-	do a	mãe	den-	tro à	en-	sea-	da
8.4.7	1	2	3	4	5	6	7	8	9	10	11
	Ex-	tin-	guir	sem	re-	mé-	dio a in-	fe-	liz	gen-	te,
8.5.4	1	2	3	4	5	6	7	8	9	10	11
	Com	que ia	a fal-	ta	pró-	pria	re-	pa-	ran-	do.	
8.11.4	1	2	3	4	5	6	7	8	9	10	11
	No-	vo	mo-	ti-	vo	pa-	ra	mais	que	rê-	lo;
8.33.2	1	2	3	4	5	6	7	8	9	10	11
	Três	naus	que em	gue-	rra	cui-	da-	do-	so ar-	ma-	ra,
8.34.7	1	2	3	4	5	6	7	8	9	10	11
	Por	bo-	ca	dos	ca-	nhões	ru-	mor	fa-	zen-	do,
8.56.4	1	2	3	4	5	6	7	8	9	10	11
	E a-	rro-	ja-	se ao	re-	bel-	de en-	fu-	re-	ci-	do.
8.60.6	1	2	3	4	5	6	7	8	9	10	11
	Des-	can-	sa-	das	do	bé-	li-	co	que	bran-	to,
8.75.1	1	2	3	4	5	6	7	8	9	10	11
	E-	ra a	di-	vi-	na	paz,	que o	céu	nos	man-	da,

**Corpus DUR: decassílabos (cont.)**

8.87.3	O-	cu-	pa o	for-	te e-	xér-	ci-	to a	ri-	bei-	ra
8.87.6	Ten-	do o	cam-	po a	Fra-	di-	que a-	co-	me-	ti-	do,
8.88.7	E	pa-	re-	ce	que	Deu-	s o	im-	pul-	so ins-	pi-
9.3.5	Pe-	tri	a	fro-	ta	que	da-	s	fn-	dias	vi-
9.19.6	No	du-	ro	jo-	go	de	Ma-	vor-	te ho-	rren-	do,
9.20.2	Con-	tra l-	ta-	ma-	ra-	cá	co-	rre o i-	ni-	mi-	go;
9.22.5	E,	fun-	dan-	do	no	céu	con-	fian-	ça	jus-	ta,
9.44.1	Ho-	lan-	da e-	ra	po-	ten-	te e o	lu-	so a-	fli-	to,
9.55.4	Con-	tra a	pra-	ça,	que au-	xí-	lio	lhe	su-	pli-	ca.
9.69.3	E,	por	mais,	que o	Vi-	ei-	ra in-	va-	da a-	ti-	vo,
9.71.3	E,	sus-	pen-	den-	do a	mi-	se-	ra	de-	rro-	ta,
10.19.3	Con-	fe-	ssa o	be-	ne-	fí-	cio a	for-	ça his-	pa-	na,
10.23.2	Don-	de	ri-	que-	za	se o-	fe-	re-	ce	tan-	ta,
10.29.2	Ta-	pa-	ri-	ca	foi	mo-	no; a	lu-	sa	gen-	te
10.42.8	Cha-	ma a	vê-	la a	pie-	do-	sa	Ca-	ta-	ri-	na.
10.60.7	E	Eu-	ro-	pa ad-	mi-	ra	pe-	lo	mar	pro-	fun-
10.63.6	Que,	an-	te-	pon-	do o	Bra-	sil	a	tu-	do a-	go-
10.69.8	Ce-	do-	to-	do o	di-	rei-	to e	en-	tre-	go a	c'ro-
											a"

**Corpus cos: hexassílabos**

105.24.4	Eu	ve-	jo (oh	do-	ri-	men-	sal)
106.17.2	Dei-	xa-	do a	pom-	pa	ver-	de,
107.1.1	Na	mar-	gem	de-	lei-	to-	sa
108.77.6	Ao	som	do	bron-	ze	ru-	do
114.6.1	Vós,	lhes	di-	go,	so-	no-	ras,
117.58.1	Re-	co-	lhi-	am-	se os	rai-	os
117.99.3	Qual	nun-	ca a	vil	fe-	re-	za
127.8.1	É	tu-	do ho-	rro-	r;	é	tu-
129.16.3	Pou-	co e	pou-	co	dou-	ran-	do
136.3.5	O	pei-	to	já	não	sen-	te
138.7.3	Um	ra-	mo	me	pre-	pa-	ra,
139.7.3	Já a-	go-	ra	não	pre-	pa-	ra;
145.2.2	Pas-	tor	mais	pe-	re-	gri-	no,

**Corpus cos: heptassílabos**

	1	2	3	4	5	6	7	8
135.14.1	Pas-	to-	ra	(se	bem	me	lem-	bra
146.2.1	Se em	mim,	pois,	se em	Po-	li-	fe-	mo
146.2.6	Bem	que	nes-	ta	co-	pia o	pin-	to,
149.2.2	E os	teus	pa-	ssos	vou	se-	guin-	do;

**Corpus cos: eneassílabos**

	1	2	3	4	5	6	7	8	9
143.5.6	Nes-	tes,	cam-	pos,	se	vê-	res-	pi-	rar!

**Corpus cos: decassílabos**

	1	2	3	4	5	6	7	8	9	10	11
7.2.1	U-	ma	fon-	te a-	qui h-	ou-	ve; eu;	não	me es-	que-	ço
11.1.1	For-	mo-	sa é	Da-	li-	a-	na; o	seu	ca-	be-	lo,
*11.4.3	Ve-	ja a	Eu-	li-	na; e en-	tão	sus-	pen-	da o	pran-	to.
28.1.3	Da-	qui	vem,	que a	mi-	nha al-	ma	de-	li-	ran-	te
38.1.1	Quan-	do,	for-	mo-	sa	Ni-	se,	di-	vi-	di-	do
	1	2	3	4	5	6	7	8	9	10	11
48.2.2	E	meu	pei-	to	sau-	do-	so,	que is-	to	vi-	a,
56.1.4	Fos-	te	do-	ce o-	ca-	são	de	meu	cui-	da-	do.
59.3.1	A-	go-	ra	tor-	no a	ver,	se a	pe-	dra	du-	ra
62.1.3	On-	de um	tem-	po os,	ga-	bões	dei-	xei	gro-	ssei-	ros
62.3.3	Que	da	ci-	da-	de o	li-	son-	jei-	ro en-	can-	to;
	1	2	3	4	5	6	7	8	9	10	11
69.2.1	Con-	fun-	da-	te a	so-	ber-	ba	ti-	ra-	ni-	a,
101.4.3	Se	per-	de	Por-	tu-	gal,	se o	mun-	do	per-	de
101.20.18	Aos	prê-	mio-	s a-	bre; a	Ge-	ne-	ral	te	cha-	ma;
101.26.19	De u-	ma al-	ma	pu-	ra e-	le	fe-	liz	vo-	ta-	va
102.1.2	E-	ssa es-	fe-	ra	de	lu-	zes,	vais	pi-	san-	do
	1	2	3	4	5	6	7	8	9	10	11
103.3.6	Ven-	tu-	ro-	so	Bai-	xel	em	gol-	fo ins-	tá-	vel
103.4.17	E	pois	te en-	con-	tro a-	go-	ra	tão	pro-	pí-	cio,
104.11.3	A-	cre-	di-	tan-	do em	di-	li-	gên-	cias	gra-	ves
104.18.2	Res-	plen-	de-	ceu	vo-	ssa	vir-	tu-	de,	pos-	ta
105.36.3	A-	qui	cho-	ran-	do es-	tou	mi-	nha	mu-	dan-	ça;

**Corpus cos: decassílabos (cont.)**

	1	2	3	4	5	6	7	8	9	10	11	
105.42.5	Se	o	deus,	que	o	tem-	pe-	ra-	do	co-	ro	te- ce,
106.29.3	O	meu	bem,	se	tal-	vez	me	não	en-	ga-	no:	
106.40.3	De	tu-	a,	mais	que	bár-	ba-	ra	im-	pie-	da-	de,
106.53.2	As	Nin-	fas	na	for-	mo-	sa	com-	pa-	nhi-	a,	
106.79.3	Oh!	Tron-	co	ge-	ne-	ro-	so!	Oh!	Plan-	ta	ra-	ra!

	1	2	3	4	5	6	7	8	9	10	11	
107.4.6	Des-	te-	rra	to-	da	a	som-	bra	do	tor-	men-	to.
108.61.2	A	pal-	ma	en-	tre	vó-	s ou-	tros;	por-	que	ve-	nho
*108.66.3	Que	eu	di-	go	o	meu,	tu,	lê	o	teu	con-	cei- to,
108.81.1	O	no-	sso	hu-	mil-	de	gê-	nio	não	con-	sen-	te
108.94.3	Tão	lon-	ge	da	ra-	zão	meu	de-	sa-	ti-	no,	

	1	2	3	4	5	6	7	8	9	10	11	
108.101.2	Nem	ga-	do,	nem	cu-	rral;	que	não	ge-	me-	sse.	
108.121.3	A	su-	a	cho-	ça	ca-	da	qual	pro-	cu-	ra:	
109.1.1	Se	é	cer-	to	que	in-	da	vi-	ve	a	do-	ce a- ve- na
110.5.1	Ah!	Que	a-	rran-	ca-	s um	mí-	se-	ro	sus-	pi-	ro
111.31.2	Só	pre-	za-	ri-	a	te-	res-	ta	cer-	te-	za,	

	1	2	3	4	5	6	7	8	9	10	11	
111.44.2	A	fru-	ta	sa-	bo-	ro-	sa	e a	ces-	ti-	nha	
114.10.4	Seu	no-	me	nes-	ta	pe-	nha	se	gra-	va-	va-	
115.1.6	Pe-	lo	man-	ce-	bo	de	me-	lhor	ju-	í-	zo,	
116.74.6	A	pro-	vi-	dên-	cia	su-	ma	no-	s en-	co-	bre.	
117.2.3	A	ver	se	dei-	xo,	pe-	la	voz	do	pran-	to,	

	1	2	3	4	5	6	7	8	9	10	11	
117.91.5	Que	na	bem	tem-	pe-	ra-	da,	a-	cor-	de a-	ve-	na,
118.2.11	E-	ssa	dor;	bem	que	a	ve-	jo	tão	pro-	fun-	da,
118.4.4	Te	fa-	zi-	a-	s ao	mon-	te	li-	son-	jei-	ro?	
118.23.3	Não	teu	co-	nhe-	ci-	men-	to,	a-	ma-	do	Al-	ga- no,
118.24.8	Da	po-	bre	cho-	ça	as	por-	tas	vão	ce-	rran-	do.

	1	2	3	4	5	6	7	8	9	10	11	
118.25.5	Mas	já	que	es-	te	su-	ce-	sso	te	per-	sua-	de
119.6.7	Rom-	pe	o	si-	lên-	cio	gra-	ve-	men-	te	mu-	do,
120.20.3	Que	A-	mor	di-	ssi-	mu-	la-	do	o	fe-	re-	ce- ra.
120.33.2	Tan-	tas	ve-	zes	nos	bai-	les,	na	flo-	res-	ta	
121.24.3	A	ver,	se	ve-	jo	a	nin-	fa,	por	quem	ge-	mo,

	1	2	3	4	5	6	7	8	9	10	11	
123.2.1	E	crês,	Lí-	ci-	da	meu,	que	sou	tão	lou-	co,	
123.9.1	Can-	ta	o	que	te	a-	gra-	dar;	mas	con-	si-	de- ra
127.1.2	Que	por	mais	ver-	de,	e	plá-	ci-	do	te-	rre-	no,
130.9.3	Do	tor-	men-	to	fa-	tal	da	vo-	ssa	au-	sên-	cia;
130.9.5	Quan-	do	do	vo-	sso	es-	pí-	ri-	to	su-	a-	ve

**Corpus cos: decassílabos (cont.)**

	1	2	3	4	5	6	7	8	9	10	11
131.5.2	A	quem	va-	lor	não	bas-	ta,	nem	es-	cu-	do,
131.21.6	Al-	ga-	no	meu,	que	re-	fe-	rir	pu-	de-	ra
131.24.6	Pa-	ra	me	não	le-	var	trás	meu	cui-	da-	do,
131.25.4	A	fú-	ria	que	meu	pei-	to	de-	bi-	li-	ta:
146.1.15	Su-	fo-	ca, es-	tra-	ga,	de-	sa-	len-	ta, e	ma-	ta;
	1	2	3	4	5	6	7	8	9	10	11
148.3.5	Que a	pe-	nha, em	que	Ni-	can-	dro	me	fa-	la-	va,
148.6.3	O	mar	tem-	pes-	tu-	o-	so; o	ven-	to i-	ra-	do
149.1.2	On-	de	te en-	con-	tra-	rei,	be-	la	Pas-	to-	ra!

**Corpus gon: decassílabos**

	1	2	3	4	5	6	7	8	9	10	11
1.1.15	Tra-	ta-	se a-	qui	da hu-	ma-	ni-	da-	de a-	fli-	ta;
2.1.2	A-	bre o-	s o-	lhos,	bo-	ce-	ja, es-	ten-	de os	bra-	ços
2.8.7	Tem	o-	lho-	s en-	co-	va-	dos,	bar-	ba	te-	sa,
2.11.57	Na-	s ou-	tras	con-	ta-	rei,	pre-	za-	do a-	mi-	go,
3.5.8	Faz	que	cho-	ra,	sus-	pi-	ra,	fe-	re o	pei-	to
	1	2	3	4	5	6	7	8	9	10	11
3.11.4	Aos	cor-	vo-	s e ao-	s a-	bu-	tres,	que	se a-	jun-	tam
4.1.47	Co-	mi-	das	da	fe-	rru-	gem?	Que a-	s al-	ge-	mas,
4.1.49	Que o	tor-	pe e-	xe-	cu-	tor	das	gra-	ves	pe-	nas
5.3.4	O	seu	ca-	cha-	ço é	gor-	do, o	ven-	tre in-	cha-	do,
5.3.24	A-	qui,	pre-	za-	do a-	mi-	go,	não	se a-	ten-	de
	1	2	3	4	5	6	7	8	9	10	11
5.3.84	Pri-	mei-	ro	lhes	pe-	diu	que	lhe	di-	sse-	ssem
5.7.23	Os	des-	pa-	chos	que	deu	no	seu	go-	ver-	no.
6.1.10	Eu	crei-	o,	Do-	ro-	teu...	Po-	rém	a-	on-	de
6.6.25	Poi-	s ain-	da,	Do-	ro-	teu,	não	vis-	te	na-	da.
6.12.4	A-	té	que	mon-	tam	a	dou-	ra-	da	se-	ge.
	1	2	3	4	5	6	7	8	9	10	11
7.2.33	Mil	coi-	sas,	Do-	ro-	teu,	mil	coi-	sas	fei-	as
7.10.3	A-	mi-	go	Do-	ro-	teu,	é	tem-	po, é	tem-	po
7.10.28	Ao	me-	no-	s u-	ma ho-	nes-	ta e	jus-	ta	cau-	sa!
7.10.35	Não	que-	res	que os	pa-	ren-	tes	se-	jam	pi-	os,
8.3.3	Que	fa-	zem	os	mi-	nis-	tro-	s: e-	le	mes-	mo



**Corpus GON: decassílabos (cont.)**

	1	2	3	4	5	6	7	8	9	10	11
8.3.14	Em	não	que-	rer	lar-	ga-	rao	vil	Ma-	tú-	sio
9.6.2	Que	po-	ssa	ser	ju-	i-	za	pró-	pria	par-	te.
9.15.9	Lhe or-	de-	na	sa-	tis-	fa-	ça	to-	do o	res-	to,
9.16.47	A-	mi-	go	Do-	ro-	teu,	se a-	ca-	so	vi-	res
10.1.11	Que	ti-	nha	nes-	ta	ca-	sa	de al-	gum	pre-	ço.
	1	2	3	4	5	6	7	8	9	10	11
10.9.10	Se	tu	a	me-	re-	ci-	a-	s? A	vi-	le-	za
10.9.31	Quais	são	os	teus	ser-	vi-	ços?	Quais?	Res-	pon-	de.
10.9.35	Te	fa-	zem,	no	con-	cei-	to	do	teu	che-	fe,
10.10.54	Se, à	for-	ça	de	ver-	go-	nha,	se	far-	da-	vam.
10.12.3	Que as	no-	ssas	leis	não	dão	aos	bons	va-	ssa-	los
	1	2	3	4	5	6	7	8	9	10	11
11.2.80	O	né-	scio	pro-	ce-	de-	r. A-	qui	não	pa-	ra
11.5.18	Que, en-	ten-	do,	nun-	ca	viu	ne-	nhu-	ma i-	da-	de!
11.5.33	O	por-	tei-	ro	ri-	so-	nho ao	li-	ci-	tan-	te,
12.2.9	Por	ci-	ú-	mes	cru-	éis,	da	sua	a-	má-	sia;
12.2.52	Que	já,	pe-	lo	ca-	mi-	nho,	vim	com-	pon-	do.
	1	2	3	4	5	6	7	8	9	10	11
12.5.10	A	quar-	to os	seus	jor-	nai-	s, e	crei-	o, a-	mi-	go,
12.15.1	Co-	mo es-	tá	be-	lo!	Co-	mo es-	tá	cas-	qui-	lho!
13.7.24	Ce-	le-	bram-	se os	fes-	tin-	s, e	não	es-	ca-	pa

**Corpus MAG: tetrassílabos**

	1	2	3	4	5
55.11.11	Em	vós	be-	le-	za,

**Corpus MAG: pentassílabos**

	1	2	3	4	5	6
3.16.4	Do	mo-	ço o	vi-	gor,	
9.21.1	Se	tu	sem-	pre a-	ten-	ta

**Corpus MAG: hexassílabos**

	1	2	3	4	5	6	7
1.3.4	Quan-	do u-	ma al-	ma i-	no-	cen-	te
9.2.3	Con-	tem-	plai	o	pa-	ssa-	do;
12.3.4	Da	ví-	ti-	ma	su-	bli-	me.
15.5.7	O	tris-	te	so-	li-	tá-	rio,
*15.7.6	É	tu-	do a	ex-	pe-	riên-	cia?

**Corpus MAG: hexassílabos (cont.)**

	1	2	3	4	5	6	7
15.13.11	A-	que-	les	que	me in-	fa-	mam;
18.21.11	Re-	ce-	be es-	ta	co-	ro-	a,
24.12.6	Re-	li-	gi-	o-	so a-	cen-	to.
24.17.3	Ri-	cas	ci-	da-	des	ven-	do.
29.10.18	Pa-	ra hon-	ra	do-	s hu-	ma-	nos,
	1	2	3	4	5	6	7
31.4.6	Pro-	vi-	den-	te	Na-	tu-	ra
41.1.27	Co-	mo a-	tra-	vés	do-	s e-	vos
46.5.5	Por	ve-	r al-	guns	teus	fi-	lhos,
50.10.8	De es-	tre-	la-	s es-	col-	ta-	dos,

**Corpus MAG: heptassílabos**

	1	2	3	4	5	6	7	8
1.24.2	Tra-	ze-	me u-	ma har-	pa	dos	céus;	
3.21.3	Tu	és	do E-	ter-	no um	sus-	pi-	ro,
10.6.2	E	de	flor-	zi-	nhas	se a-	rrei-	am;
10.10.1	Tu-	do	vi-	da in-	da	res-	pi-	ra;
17.4.2	Pa-	ra a	ten-	ra	cri-	a-	tu-	ra,

**Corpus MAG: decassílabos**

	1	2	3	4	5	6	7	8		
29.3.7	Quem	po-	r o-	lhos	so-	be-	ra-	nos		
36.4.4	A-	ssim	po-	de	sus-	pi-	rar,			
43.6.3	E o	as-	tro	que	me a-	cla-	ra-	va		
44.5.4	Um	a-	cen-	to	de	tris-	te-	za.		
56.6.3	Vol-	to a	ver	te-	rras	da	Pá-	tria,		

	1	2	3	4	5	6	7	8	9	10	11
2.3.4	Tu	no-	s ho-	mens	só	vês	vir-	tu-	de, ou	ví-	cio.
2.4.4	Teu	íg-	neo, a-	la-	do	gê-	nio,	no ar	sus-	pen-	so:
3.7.2	Lá	no	mei-	o	da	noi-	te,	quan-	do a	lu-	a
4.10.2	Pa-	ra im-	pe-	dir-	lhe a	mar-	cha, as	fron-	te-	s er-	guem
4.17.1	Eu	te	ve-	ne-	ro, oh	Deus	da Hu-	ma-	ni-	da-	de!

	1	2	3	4	5	6	7	8	9	10	11
6.11.5	Di-	vi-	na ins-	pi-	ra-	ção,	tu	só	po-	di-	as
6.11.8	Nem	sem-	pre	ver-	da-	dei-	ros.	Tu	re-	ve-	las
9.8.4	Se	vo-	s a-	pra-	z o	bem,	se o	bem	vos	cha-	ma.
9.13.4	Que	gi-	ran-	do	no	cen-	tro	do U-	ni-	ver-	so,
10.4.5	Da	vi-	da	vo-	s a-	pa-	ga a	tê-	nue	fla-	ma.

**Corpus MAG: decassílabos (cont.)**

	1	2	3	4	5	6	7	8	9	10	11
12.3.2	Sus-	pi-	ra a	ca-	da	go-	ta	de-	sse	san-	gue,
13.3.1	A-	qui	so-	bre es-	ta	pe-	dra	so-	li-	tá-	rio
13.10.3	Com a	mes-	ma	ra-	pi-	dez	com	que	num	so-	nho
13.15.5	Com o	san-	gue	de	Lu-	ís,	de um	rei	com o	san-	gue.
13.17.1	Já	não	ve-	des,	meu-	s o-	lhos;	no-	vas	tre-	vas
	1	2	3	4	5	6	7	8	9	10	11
16.15.6	Mes-	mo	quan-	do os	ru-	ins	do-	res	lhe	cau-	sem,
16.17.5	Mas	de-	pois	da	tre-	men-	da	tem	pes-	ta-	de,
18.18.1	O	sol	des-	maia;	e	pre-	cur-	sor	da	noi-	te
19.9.12	On-	de	sem	ris-	co	to-	dos	nós	mar-	cha-	mos,
20.3.2	Mãe	i-	lus-	tre	de he-	róis	do	mun-	do es-	pan-	to!
	1	2	3	4	5	6	7	8	9	10	11
20.4.2	Qu'in-	da e-	xa-	la	mor-	tí-	fe-	ros	va-	po-	res;
21.4.5	Co-	mo ao	la-	do	do al-	tar	pen-	den-	te	vo-	to
21.9.3	Teus	ri-	o-	s? e o	teu	céu	a-	zul	sem	nó-	doa?
22.24.12	Oh	Ra-	zão,	o	que és	tu? —	Ím-	pios,	ca-	lai-	vos,
24.5.7	Tão	lon-	ges	prai-	as! —	Quem	me	de-	ra eu	vê-	las!
	1	2	3	4	5	6	7	8	9	10	11
24.16.2	Que íg-	neos	rai-	os	re-	fle-	tem,	si-	mu-	lan-	do
24.17.13	In-	da a	seus	pés	ve-	rás	mur-	chas	sau-	da-	des,
27.2.2	Cir-	cu-	la-	dos	de au-	réo-	las	mo-	ri-	bun-	das,
29.9.7	Que o	ú-	ni-	co	não	fos-	te,	que	da	sor-	te
44.18.4	Mi-	nha al-	ma	se	di-	la-	ta, e a-	ro-	mas	ver-	te;
	1	2	3	4	5	6	7	8	9	10	11
47.7.14	A-	ssaz	não	ti-	nha o	Mun-	do	co-	nhe-	ci-	do,
51.4.1	Es-	tu-	da,	sim	es-	tu-	da,	mas	pra-	ti-	ca
54.11.4	De	te	ver,	de	te ou-	vi-	r, e	mai-	s a-	in-	da,
55.3.18	Tan-	tas	ve-	ze-	s a-	cha-	mo-	s o	des-	can-	so
55.4.9	A-	gro	pra-	zer	de um	co-	ra-	ção	ma-	goa-	do,

**Corpus PIN: decassílabos**

	1	2	3	4	5	6	7	8	9	10	11
2.7.2	Que	pa-	ra	Ro-	ma e o im-	pé-	rio	se	fun-	da-	rem
2.19.3	De an-	jo	com	voz,	fa-	lan-	do-	me	pie-	do-	sa:
2.20.1	—“De	Mân-	tua al-	ma	cor-	tês,	que in-	da ho-	je em	di-	a
3.39.3	Qual	a-	ve	dos	re-	cla-	mo-	s a-	tra-	í-	da.
4.8.3	Em	que o a-	bis-	mo a es-	trei-	tar-	se	já	co-	me-	ça,

**Corpus PIN: decassílabos (cont.)**

4.29.3	Em-	pu-	nha es-	pa-	da e os	três	dei-	xa	dis-	tan-	tes.
4.37.2	En-	trei	por	se-	te	por-	tas,	ca-	mi-	nhan-	do
6.17.3	E-	xis-	tên-	cia	go-	zei	le-	da e	se-	re-	na.
7.26.3	Ge-	ral	mi-	nis-	tra e	di-	re-	to-	ra	deu-	lhes,
9.13.1	Lá	o as-	pec-	to	se er-	guia h-	o-	rri-	pi-	lan-	te
9.19.1	—“Vol-	ta a	fa-	ce;	de	lu-	z o	ros-	to	fal-	to
9.26.1	Quais	rãs,	que	di-	vi-	san-	do a	co-	bra i-	mi-	ga,
9.43.3	São	mais;	do	que	tu	crê-	s em	ca-	da en-	te-	rro.
10.1.2	Que en-	tre o	mu-	ro e os	mar-	tí-	rios	vai	se-	guin-	do:
*12.37.3	O-	biz-	zio	d’Es-	te, ao	qual	(ver-	da-	de eu	di-	go)
13.7.1	“A-	té	que ao	a-	re-	al	che-	gue-	s in-	fan-	do.
14.25.2	Não	ca-	mi-	nhes;	da a-	reia	in-	can-	de-	scen-	te;
14.29.2	De-	pois	que a-	trás	dei-	xa-	mo-	s e-	ssa	por-	ta,
14.32.3	Com	seu	rei	foi	do	mal	o	mun-	do i-	sen-	to.
14.43.2	Não	re-	ve-	le o	teu	ros-	to	ma-	ra-	vi-	lha.
14.47.1	Di-	sse	de-	poi-	s: “É	tem-	po	de	dei-	xar-	se
15.16.3	De	quem	re-	ce-	bes	na	jor-	na-	da en-	si-	no?”
17.5.3	Mil	ro-	de-	la-	s e	nós	se en-	tre-	la-	çan-	do.
17.6.2	Tár-	ta-	ros,	Tur-	cos	nun-	ca	mis-	tu-	ra-	ram,
19.41.3	Ver-	da-	des;	que eu,	sin-	ce-	ro,	pro-	fe-	ri-	ra.
20.16.1	“Vol-	ta-	lhe A-	ron-	s ao	ven-	tre as	cos-	tas	nu-	as:
21.46.3	Por	mo-	fa a	lín-	gua ao	ca-	bo	já	dis-	pos-	to;
22.5.3	U-	rrar	com	lo-	bos	dis-	cri-	ção	re-	ve-	la.
23.3.3	Quan-	do o	prin-	cí-	pio e	fim	seus	re-	cor-	da-	va.
23.38.1	Em	ven-	do-	me es-	tor-	ceu-	se an-	gus-	ti-	a-	do.
26.26.3	Nes-	ta	lin-	gua-	gem	di-	sse	per-	sua-	si-	va:
27.2.3	O-	lhar	nos	fez	seu	ci-	mo,	que o	sci-	la-	va.
27.11.1	Pa-	ra es-	cu-	ta-	r a-	ten-	to a	fron-	te in-	cli-	no.
27.31.1	“Sa-	cra-	s or-	den-	s em	si	não	con-	si-	de-	ra,
28.5.1	E a	que	fo-	ra	por	gol-	pe-	s a-	ba-	ti-	da,
28.11.3	Do	men-	to à	tes-	ta o	ros-	to	lhe é;	des-	fei-	to.
28.36.1	Gri-	tou: —“	Me-	mo-	ra	Mos-	ca	des-	di-	to-	so!
29.32.1	—“Eu	sou	que”—	diz	Vir-	gí-	lio ao	des-	di-	to-	so
29.37.2	Mor-	te	me	deu	nas	cha-	mas,	tru-	cu-	len-	to,
*30.10.1	U-	ma al-	can-	çou	Ca-	poc-	chio e	lhe	cra-	van-	do

**Corpus PIN: decassílabos (cont.)**

32.18.3	1	2	3	4	5	6	7	8	9	10	11
	"Por	que	teu-	s o-	lhos	sô-	fre-	gos	nos	mi-	ram?
32.35.2	Da	gue-	de-	lha u-	ma	par-	te a-	rre-	pe-	la-	ra:
33.12.1	"Em	bre-	ve o	pai	e os	fi-	lhos,	la-	ssos,	man-	cos,
33.32.2	E a	dor,	que	de-	sa-	fo-	go em	vão	pro-	cu-	ra,
34.36.3	Do	que a-	tra-	ve-	ssa o	mun-	do,	ver-	me in-	fes-	to.
35.5.1	1	2	3	4	5	6	7	8	9	10	11
	Su-	a-	ve	cor	de o-	ri-	en-	tal	sa-	fi-	ra,
36.3.1	Já	ne-	sse	tem-	po a i-	da-	de	trans-	for-	ma-	va
36.21.1	Tor-	nou	Vir-	gí-	lio: —	Vo-	s a-	praz	jul-	gar-	nos
37.18.2	"Quem	sa-	be on-	de a es-	te	mon-	te o	pa-	sso a-	scen-	de?
39.19.1	"A-	ssim,	con-	tri-	tos,	per-	do-	an-	do,	fo-	ra
39.25.3	1	2	3	4	5	6	7	8	9	10	11
	De An-	te-	nó-	rios	nos	pla-	nos	pan-	ta-	no-	sos,
39.35.1	"Di-	rei	ver-	da-	de: aos	vi-	vo-	s o a-	nun-	ci-	a.
44.3.3	Qual	on-	da,	que o-	ra a-	cer-	ca, o-	ra	des-	vi-	a.
44.45.2	Em	quem	o-	lha:	sen-	ti	en-	tão	no-	tan-	do
45.40.3	E-	sse,	que a-	pon-	tas,	con-	ta-	me	quem	e-	ra." —
46.6.1	1	2	3	4	5	6	7	8	9	10	11
	A-	ssim	co-	mo	na	cam-	pa	se a-	vi-	ven-	ta
46.14.2	So-	bre a	mon-	ta-	nha	Gel-	bo-	é —	mal-	di-	ta,
46.46.1	So-	rriu-	se o	Mes-	tre,	que o	meu	ges-	to	vi-	ra.
47.4.2	Que	nos	pres-	te	con-	se-	lho" —	diz	meu	Gui-	a —
48.13.3	Do	cli-	ma, ou	por	mo-	ral,	que o	bem	re-	fu-	ga.
49.42.1	1	2	3	4	5	6	7	8	9	10	11
	— "Vou	ex-	por,	meu	bom	mes-	tre, aos	teu-	s ou-	vi-	dos" —
49.47.1	Na-	s ho-	ras	ves-	per-	ti-	nas	pre-	ssu-	ro-	sos
51.12.1	Don-	ze-	la eis	na	vi-	são	se	me	de-	pa-	ra
51.12.3	Por	que	to-	mas-	te i-	ro-	sa a	mor-	te a-	ma-	ra?
51.36.1	"E	co-	mo o a-	mo-	r o	bem	so-	men-	te es-	tu-	de
51.38.3	1	2	3	4	5	6	7	8	9	10	11
	No hu-	ma-	no	li-	mo em	mo-	dos	trê-	s a-	tu-	a.
52.6.2	No es-	pi-	ri-	to e há	de	ser-	te	bem	pa-	ten-	te
53.30.1	Da	per-	mi-	ssão	do	Mes-	tre u-	san-	do	pres-	to,
54.6.3	Por	car-	pir	la-	men-	to-	so a	dor	sen-	ti-	da.
54.13.2	Le-	var-	me-	de-	ve ao	sus-	pi-	ra-	do	por-	to,
54.22.2	1	2	3	4	5	6	7	8	9	10	11
	Co-	me-	çan-	do, in-	va-	diu	por	pe-	ni-	tên-	cia
54.25.2	Lan-	ça	de	Ju-	das,	que a	Flo-	ren-	ça a-	pon-	ta:
55.28.3	San-	gue,	por	Ju-	das" —	re-	pli-	cou —	ven-	di-	do,
56.22.3	E	de	Deus	san-	ta	luz	ver	me	fi-	zes-	te.
56.46.2	Da	par-	te, em	que a	mon-	ta-	nha a en-	tra-	da	mu-	ra;

**Corpus PIN: decassílabos (cont.)**

	1	2	3	4	5	6	7	8	9	10	11
57.22.1	"Os	que o	ros-	to,	can-	tan-	do,	têm-	ba-	nha-	do
57.30.2	A-	ssim,	me al-	cei	da en-	cos-	ta, on-	de	se es-	pe-	ra,
57.36.2	Do	cas-	ti-	go,	que es-	tá	lhe-	s i-	mi-	nen-	te,
58.43.2	E as	pe-	na-	s es-	cu-	ta-	mos	dos	pe-	ca-	dos
61.27.3	Com	seu	ca-	ja-	do e o-	s o-	lhos	pre-	vi-	den-	tes;

	1	2	3	4	5	6	7	8	9	10	11
64.5.3	Com	re-	no-	va-	da	vo-	z a-	le-	lui-	an-	do,
64.30.1	De-	pois	li-	que	sce e	co-	rre	de-	rre-	ti-	da
64.32.1	Mas	quan-	do, em	seus	con-	cen-	to-	s ex-	pre-	ssa-	da,
66.44.1	A	te-	rra en-	tão	me	pa-	re-	ceu	se a-	bri-	a,
66.53.1	No	ci-	ú-	me e	na	i-	ra,	que o in-	fla-	ma-	va

	1	2	3	4	5	6	7	8	9	10	11
69.23.3	In-	flu-	in-	do	com	mais,	me-	nos	gra-	ve-	za:
70.25.2	Se-	rí-	a-	mos	re-	bel-	de-	s à	von-	ta-	de,
71.15.3	Pê-	s e	mão-	s ao	Se-	nhor	do	fir-	ma-	men-	to.
71.25.2	A	ví-	ti-	ma	se o-	pon-	do, em	lu-	ta in-	sis-	te,
71.27.2	Ce-	de à	for-	ça:	vol-	ta-	rao	san-	to a-	bri-	go

	1	2	3	4	5	6	7	8	9	10	11
71.35.1	"Do	pai	ar-	den-	tes	ro-	gos	res-	pei-	tan-	do,
72.30.2	Ca-	lou-	se o	meu	de-	se-	jo im-	pa-	ci-	en-	te,
73.40.1	"Mas	no-	ssa	di-	ta em	par-	te	com-	pre-	en-	de
74.3.3	Ve-	lo-	zes,	co-	mo	fla-	ma	ful-	gu-	ro-	sa.
74.17.3	Vin-	ga-	da	ser	de-	vi-	a	por	se-	gun-	da.

	1	2	3	4	5	6	7	8	9	10	11
74.48.1	"A	vi-	da	no-	ssa	vem	di-	re-	ta-	men-	te
75.1.3	Lou-	co a-	mor	do e-	pi-	ci-	clo	que é	ter-	cei-	ro.
75.4.2	Ao	pla-	ne-	ta,	que o	sol	sem-	pre	na-	mo-	ra,
75.6.3	Fir-	me o	som	de u-	ma, o	de ou-	tra	va-	ri-	a-	do,
75.14.3	E	cer-	to	fui	do	seu	que-	rer	do-	no-	so,

	1	2	3	4	5	6	7	8	9	10	11
75.37.1	"Fo-	ram	en-	tão	de	per-	fei-	ção	des-	pi-	das
76.19.2	Que o	san-	gue	re-	ce-	be-	sse	de	Fe-	rra-	ra,
77.47.3	Ma-	ti-	nas	can-	ta e o	seu	a-	mo-	r im-	plo-	ra;
78.6.3	Do	seu	bri-	lho a	pu-	re-	za	se au-	men-	ta-	ra:
78.13.2	Ou-	tro os-	ten-	tou,	por	seu	sa-	ber	na	te-	rra,

	1	2	3	4	5	6	7	8	9	10	11
78.25.1	"Pa-	ra	fa-	zer	mi-	nha	lin-	gua-	gem	cla-	ra,
79.48.3	O	di-	zer	sá-	bio	de	To-	más	de A-	qui-	no,
81.36.1	Quem	to-	ma a	Cru-	ze	na	jor-	na-	da	Cris-	to
82.2.3	Que	lá	do	céu	a	des-	tra a-	frou-	xa e	es-	ti-
82.12.1	No-	s o-	lhos	seu-	s ar-	dia	um	tal	so-	rri-	so,

**Corpus PIN: decassílabos (cont.)**

84.19.1	1	2	3	4	5	6	7	8	9	10	11
	"Dei-	xa-	rás	to-	da a	cau-	sa a	mais	que-	ri-	da,
84.27.2	Só	vol-	tas	no-	ve em	tor-	no	lhe	tem	da-	do
85.39.3	Do	céu	de	que é	s or-	na-	to	per-	ma-	nen-	te!
85.43.1	Com	glá-	dio ou-	tro-	ra	se	tra-	va-	a	gue-	rra;
86.16.3	Não	es-	pe-	ran-	do a	luz	ca-	iu	a-	cer-	bo.
87.20.3	Pos-	to ha-	ja o	mun-	do ho-	rren-	do	des-	pro-	vei-	to.
87.39.1	"E	foi	na	fé,	no a-	mor	tão	fer-	vo-	ro-	sa,
87.44.1	"Ó	pre-	des-	ti-	na-	ção!	Não	te al-	can-	ça-	ram
88.8.3	A-	pó-	s u-	ma	go-	zan-	do ou-	tra a-	le-	gri-	a.
88.10.1	De áu-	rea	co-	r, em	que o	sol	é	re-	fle-	ti-	do,
88.20.1	"E	por	que	mu-	da es-	tá	na es-	fe-	ra	cla-	ra
88.20.3	Que	tão	de-	vo-	ta	nou-	tra-	s es-	cu-	ta-	ra." —
88.33.3	Os	pés	mo-	ven-	do ao	tran-	sce-	den-	te	fi-	to.
89.18.2	Tor-	nei —	"e o	bem	que-	rer,	que	tão	pa-	ten-	te
89.27.3	Que aos,	mon-	ges	tem	os	co-	ra-	çõe-	s ei-	va-	do;
89.39.3	Quan-	do as-	pi-	rei	pri-	mei-	ro ar	da	Tos-	ca-	na.
89.41.3	Ár-	duo	pa-	sso	que a	si	me a-	traí,	me	pren-	de.
90.22.3	A	mal	não	le-	va-	rá,	se ao	car-	go eu	tre-	ma.
92.31.1	"Di-	sse l-	sa-	í-	as:	ca-	da	qual	cin-	gi-	da
92.35.2	E,	pa-	ra a	noi-	va hon-	rar,	dan-	ça i-	no-	cen-	te
92.38.2	O	no-	sso	Pe-	li-	ca-	no e	de-	le há	si-	do
93.33.2	Mo-	vi-	men-	to,	que os	mem-	bros	seu-	s a-	gi-	ta
93.34.1	A-	ssim	e-	ssa	pri-	mei-	ra al-	ma	ben-	di-	ta
93.35.2	"Teu	de-	se-	jo,	me-	lhor	que	tu	sa-	ben-	do
95.23.2	De	mor	bem	foi	mor	cor-	po a-	qui-	nho-	a-	do,
96.4.2	A-	go-	ra	não	te in-	qui-	ro:	já	o hei	vis-	to
96.18.3	Des-	te e	xcel-	so	mis-	ter	se	não	dis-	par-	te.
97.7.2	Não	se-	ria	o	pri-	mor,	que	vi	mas,	crei-	o,
97.10.1	Dês	que	vi	do	seu	ges-	to o	pa-	ra-	í-	so
98.5.2	Na-	s a-	sa-	s ou-	ro, em	tu-	do	mai-	s al-	vu-	ra,
98.7.3	Nem	vis-	ta,	nem	splen-	do-	re-	s a-	ta-	lha-	vam,

**Corpus ANJ: hexassílabos**

	1	2	3	4	5	6	7
78.2.3	Da	qual	bru-	ta,	pro-	vi-	nha
158.2.1	Teus	sei-	o-	s, oh!	mo-	re-	na
158.4.4	Dos	bra-	ços	teus	na	Cruz!	

**Corpus ANJ: heptassílabos**

	1	2	3	4	5	6	7	8
281.2.4	Ei	pe-	ru-	de-	ro-	da o-	be-	sa.

**Corpus ANJ: decassílabos**

	1	2	3	4	5	6	7	8	9	10	11
1.12.6	Do-	s a-	po-	dre-	ci-	men-	tos	mus-	cu-	la-	res!
6.1.3	Dei-	xa	cir-	cun-	fe-	rên-	cias	de	pe-	ço-	nha,
6.1.4	Mar-	ca-	s o-	riun-	das	de úl-	ce-	ra-	s e an-	tra-	zes.
8.2.4	A	mi-	nha	mor-	fo-	gê-	ne-	se an-	ces-	tral?!	
9.1.4	Pa-	ra	la-	tir	nas	so-	li-	dõe-	s e-	nor-	mes?!

	1	2	3	4	5	6	7	8	9	10	11
12.25.4	A he-	mo-	glo-	bi-	na	vi-	nha	chei-	a	de á-	gua!
13.1.3	Que im-	por-	ta a	mim	que a	bi-	cha-	ri-	a	ro-	a
13.2.4	Ao	con-	ta-	to	de	bron-	ca	des-	tra	for-	te!
21.2.1	A	di-	ges-	tão	de-	sse	man-	jar	fu-	né-	reo
25.17.4	Vô-	mi-	to-	s im-	preg-	na-	dos	de	ptia-	li-	na.

	1	2	3	4	5	6	7	8	9	10	11
*25.19.2	A-	di-	vi-	nhan-	do o	fri-	o	que há	nas	lou-	sas,
25.20.3	Es-	tas	dos	ba-	ca-	lhau-	s o ó-	leo	não	be-	bem,
25.49.3	Mal-	di-	zi-	am,	com	vo-	z es-	ten-	to-	ro-	sa,
25.60.2	À	subs-	tân-	cia i-	nor-	gã-	ni-	ca	pri-	me-	va
25.82.3	On-	de	mi-	nh'al-	ma in-	tei-	ra	sur-	preen-	di-	a

	1	2	3	4	5	6	7	8	9	10	11
39.2.2	O	sis-	te-	ma	ner-	vo-	so	de um	gi-	gan-	te
43.7.1	E eu	sa-	í	a	tre-	mer	com a	lín-	gua	gro-	ssa
43.16.1	O	mo-	tor	te-	leo-	lô-	gi-	co	da	Vi-	da
48.7.6	Foi	so-	bre e-	ssa i-	lha	que ex-	tin-	gui	meu-	so-	nho!
50.15.6	Es-	tão	di-	zen-	do	que	Je-	su-	s é	mor-	to!

	1	2	3	4	5	6	7	8	9	10	11
50.20.3	Te-	nho o-	s o-	lho-	s em	lá-	gri-	ma-	s i-	mer-	sos...
52.12.1	Ah!	Por	to-	do-	s os	sé-	cu-	los	vin-	dou-	ros
53.2.1	Que	vo-	z é	es-	ta	que a	ge-	mer	con-	cen-	tro
53.16.4	No	meu	e-	ter-	no	lei-	to	de	Pro-	cus-	to!
56.2.1	Dois	são,	por-	que um,	cer-	to,	é	do	so-	nho a	ssí- duo



**Corpus ANJ: decassílabos (cont.)**

61.3.3	Que a-	s har-	mo-	ni-	as	si-	de-	rai-	s in-	va-	de...
67.5.5	A es-	cu-	ri-	dão	dos	gi-	ne-	ceus	fa-	li-	dos
67.11.6	De	não	ter	si-	do	fiel	ao	seu	des-	ti-	no!
70.4.2	Mi-	nha	mai-	or	ven-	tu-	ra é es-	tar	de	po-	sse
74.2.1	Em	a-	lu-	ci-	na-	tó-	rias	ca-	val-	ga-	das,
93.2.3	Pa-	ra-	li-	sa-	da e es-	tran-	gu-	la-	da,	mal	
98.11.2	A abs-	ti-	nên-	cia e a	lu-	xú-	ria, o	bem	e o	mal	
98.23.6	A es-	fin-	ge	do	Mis-	té-	rio U-	ni-	ver-	sall	
105.3.2	Se es-	ta	dú-	vi-	da	cruel	qual	me	ma-	go-	a
110.3.3	Es-	con-	de a	Na-	tu-	re-	za o	so-	fri-	men-	to,
113.1.2	A	pi-	ra	da	vin-	gan-	ça	re-	pri-	mi-	da,
122.1.4	E a	te-	rra	cho-	va	no	teu sei-	o	flo-	res!	
129.4.3	Ao	do-	bre	fu-	ne-	ral	dos tem-	po-	s i-	dos!	
143.1.4	De on-	das	cho-	ran-	do	num	al-	bo-	re-	té-	reo.
150.4.3	— Que-	ro	co-	rre-	r em	bus-	ca	do	Fu-	tu-	ro.
163.2.4	Cho-	ra	sau-	da-	des	do	te-	rre-	no	pou-	so!
165.2.3	No	si-	lên-	cio	noc-	tí-	va-	go	da	ru-	a...
168.1.1	Ân-	sias	que	pun-	gem,	mór-	bi-	do-	s en-	can-	tos,
169.1.4	Fri-	a	co-	mo um	cre-	pús-	c'lo	da	Ju-	dei-	a.
177.2.1	Ha-	los	de	luz	de	bran-	cas	mu-	sse-	li-	nas
181.1.3	En-	ches	de	bei-	jos	de in-	fi-	ni-	ta	cal-	ma...
181.7.2	Que	na	fi-	bra	mais	for-	te e a-	té	na	fi-	bra
185.2.1	Sai-	o	de	ca-	sa. Os	pa-	ssos	mal	se-	gu-	ros
189.1.1	Cha-	ma-	se	Do-	r, e	quan-	do	pa-	ssa, en-	lu-	ta
199.17.1	Es-	tes	ver-	sos	de a-	mor	que a-	go-	ra	fin-	do
219.1.3	Co-	mo u-	ma	lo-	ba he-	dion-	da	que es-	ca-	bu-	ja
241.1.2	Tra-	ta	to-	da-	s as	mo-	ças	com	ca-	ri-	nho
255.1.4	Com o o-	lhar!	— Teu-	s o-	lhos	são	de	Ba-	si-	lis-	co...
255.4.4	Com as	má-	gi-	cas	vir-	tu-	des	do	teu	Cés-	to...
256.3.2	Rou-	pa,	jul-	gan-	do	com	pra-	zer	pro-	fun-	do
271.3.1	E	que,	bem	co-	mo a á-	gua	de um	la-	go i-	mun-	do
272.4.2	A-	chin-	ca-	lhan-	do o	sen-	ti-	men-	to hu-	ma-	no
275.2.4	E um	ba-	ru-	lho	de	pé-	ro-	las	ro-	lan-	do.

**Corpus ANJ: dodecassílabos**

	1	2	3	4	5	6	7	8	9	10	11	12	13
187.1.1	A-	go-	ni-	a	de a-	ma-	r, a-	go-	ni-	a	ben-	di-	ta!
217.1.4	Sei	vi-	brar	to-	da a es-	ca-	la hie-	rár-	qui-	ca	do	som,	
217.1.44	Dei-	xa	na	car-	na-	ção	mor-	tal	de	ca-	da	ser,	

**Corpus FIG: tetrassílabos**

	1	2	3	4	5	6
*2.28.2.39	O	ce-	go,	em	êx-	ta-se:

**Corpus FIG: pentassílabos**

	1	2	3	4	5	6
2.6.6.7	Os	meus	pe-	sa-	de-	los!
2.28.2.73	Sem	con-	so-	la-	ção,	

**Corpus FIG: hexassílabos**

	1	2	3	4	5	6
2.27.3.1	— Por-	que	não	po-	sso an-	dar;

**Corpus FIG: heptassílabos**

	1	2	3	4	5	6	7	8
2.1.19.2	Pe-	las	prai-	as	per-	fu-	ma-	das.
2.1.23.1	Tal	qual	teu	man-	to	sa-	gra-	do
2.3.4.1	Pe-	los	bos-	ques	so-	li-	tá-	rios
2.3.9.2	Cha-	ma-	lo-	te-	s e	mi-	ra-	gens.
2.10.5.3	Nos	cir-	cos	de	ca-	va-	li-	nhos;
	1	2	3	4	5	6	7	8
2.10.8.3	De	lhe	dar	fre-	men-	tes	bei-	jos,
2.13.6.2	Em	a-	le-	gri-	a-	s, e	for-	mem
2.23.6.2	No	mo-	men-	to em	que	par-	tis-	te,
2.24.3.2	Ao	re-	ce-	be-	ros	con-	se-	lhos
2.24.14.1	Po-	r on-	de	pa-	ssas	des-	cal-	ça
	1	2	3	4	5	6	7	8
2.24.22.1	Quan-	do	de	mim	es-	tás	lon-	ge,
2.26.8.6	E	que	fres-	cu-	ra	nas	li-	nhas
4.1.52.1	E	tão	ba-	nha-	das	de a-	fa-	gos
4.1.73.1	Tu,	ten-	dal	a-	ber-	to em	flo-	res,

**Corpus FIG: eneassílabos**

	1	2	3	4	5	6	7	8	9	10
2.12.5.5	A	vi-	da é	a-	ssim,	de-	ssa	ma-	nei-	ra:
2.12.17.2	Des-	ta ou	tra	noi-	va,	tam-	bém	ves	ti-	da

**Corpus FIG: decassílabos**

	1	2	3	4	5	6	7	8	9	10	11
3.1.4.2	Sem	tor-	men-	tos,	so-	lu-	ço-	s e	can-	sa-	ços...
3.2.3.1	E eu	di-	sse ao-	s an-	jos: —	Pa-	ra a	gran-	de bo-	da	
3.7.1.4	Em	con-	tí-	nuas,	pro-	fun-	da	co-	rren-	te-	zas...
3.16.2.3	Não	trou-	xes-	te	no o-	lhar	to-	da a a-	le-	gri-	a?
3.17.4.3	Mai-	or	que	te-	mos	nes-	te a-	troz	ca-	mi-	nho!

	1	2	3	4	5	6	7	8	9	10	11
3.22.2.3	Não	i-	as	tu	em	bus-	ca	de al-	gum	ni-	nho?
3.24.2.1	E a	gen-	te	vê	as	fa-	ces	se en-	ru-	gan-	do,
4.6.1.2	Cris-	ta-	li-	na-	s es-	tre-	las	mis-	te-	rio-	sas!
4.6.2.1	Re-	fú-	gios	que	mi-	nha al-	ma,	de en-	tre as	du-	ras
4.16.4.1	E o a-	mor	lhe	di-	sse, a-	brin-	do-	lhe o	re-	ga-	ço!

	1	2	3	4	5	6	7	8	9	10	11	
4.22.2.4	E	sim	a	tre-	va au-	gú-	rea	que a-	li-	men-	tas!	
4.30.1.1	Ah!	quan-	do o	si-	no, e-	mo-	cio-	nal,	bim-	ba-	lha:	
4.32.2.3	Em	si	re-	co-	lha a	dor,	co-	mo um	pa-	lha-	ço;	
4.62.1.3	Vê-	lo	de	no-	vo. En-	tão,	da e-	xcel-	sa al-	tu-	ra,	
4.65.1.1	Má-	goal	E-	qui-	mo-	se a-	troz!	Quem	nun-	ca a	te-	ve?

	1	2	3	4	5	6	7	8	9	10	11
4.70.23.2	En-	quan-	to an-	dais	da	vi-	da	pe-	la es-	tra-	da,
5.5.4.3	Ve-	rei	a	á-	gua	trans-	for-	ma-	da em	vi-	nho.
5.13.3.3	Se	não	res-	pon-	de, in-	gra-	to, aos	teus	des-	ve-	los?
5.22.3.2	E	não	os	po-	de-	rás	cha-	mar	de	lou-	cos,
5.67.4.2	De	lem-	bran-	ças	do	seu	a-	mo-	r in-	ten-	so,

**Corpus FIG: dodecassílabos**

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.1.16.3	So-	lu-	ce, em-	bo-	ra, a	voz	do	ca-	rri-	lhão	da-	s ân-	sias!
1.24.4.2	De-	scem	to-	do-	s ao	ma-	re	vão	co-	lhe-	ra	me-	sse
1.30.1.2	E, ao	te	sen-	tir,	re-	cor-	do o:	teu	se-	re-	no	vul-	to,
1.35.1.2	No en-	re-	ge-	la-	do	pei-	to e	mús-	cu-	los	do Ar-	man-	do!
1.35.2.3	Pa-	re-	ci-	am	i-	guai-	s às	de-	so-	la-	das	flo-	res

**Corpus FIG: dodecassilabos (cont.)**

1.40.1.1	1	2	3	4	5	6	7	8	9	10	11	12	13
	Que	mais	lin-	da	mu-	lher	po-	ssui	o en-	can-	to e a	gra-	ça
1.42.3.3	Bi-	ca-	ra em-	bar-	ca-	ção	de al-	tos	bor-	dos	de	ce-	dro,
1.58.4.2	À	lem-	bran-	ça	me	vêm	os	sei-	os	cor	de o-	pa-	la
1.68.3.3	Ri-	ta	pôs-	se a	cis-	mar	na	vi-	da	do	Lou-	ren-	ço.
1.77.2.4	A-	bre a	ca-	da	mo-	men-	to. E	ca-	da	va-	ga é um	po-	ço.
1.78.3.3	1	2	3	4	5	6	7	8	9	10	11	12	13
	E	se-	jam	mi-	nha un-	ção	os	per-	fu-	me-	s a-	gres-	tes...
1.83.4.3	Pa-	ra	ce-	do	dor-	mir	na	paz	do	cam-	po	san-	t".
1.85.2.2	Nu-	ma	mis-	te-	ri-	o-	sa e	pu-	rí-	ssi-	ma a-	de-	ga:
1.92.1.4	Dei-	xan-	do	pa-	ra	trás,	na	sa-	fra, os	com-	pa-	nhei-	ros.
1.93.1.1	O-	lho o	ma-	r, o-	lho o	ri-	o, o-	lho o	cam-	po, o-	lho a	se-	rra,
1.95.4.1	1	2	3	4	5	6	7	8	9	10	11	12	13
	Des-	pren-	di-	dos,	de-	pois,	de	sam-	bu-	rá	às	cos-	tas,
1.98.4.3	Bus-	ca o	con-	so-	lo e a	paz	pa-	ra	to-	da-	s a-	s al-	mas.
1.106.1.4	Sem	lam-	pe-	jos	de	sol	ou	a-	fa-	gos	de	luar.	
1.108.4.3	Sa-	í-	a	pa-	ra o	ma-	r, à	pes-	ca	das	ta-	i-	nhas...
1.115.1.1	Ve-	jo,	se-	gui-	da-	men-	te, es-	plên-	di-	das	pai-	sa-	gens
1.118.1.1	1	2	3	4	5	6	7	8	9	10	11	12	13
	Que	be-	lo es-	ta-	va o	céu,	na	tar-	de i-	ma-	cu-	la-	da
1.127.3.2	Quan-	tas	re-	cor-	da-	ção-	s! E à	noi-	te, às	gran-	des	man-	tas
1.128.1.1	Se-	tem-	bro. O	céu	es-	ba-	te uns	vio-	lã-	ceos	ful-	go-	res
1.141.2.4	To-	da a	ma-	ssa	do	mar,	que	se en-	ros-	ca,	bra-	min-	do...
1.144.1.4	O	cam-	po, eu	co-	me-	ça-	va a	do-	lo-	ro-	sa	li-	da...
1.161.4.2	1	2	3	4	5	6	7	8	9	10	11	12	13
	Pe-	lo o-	ce-	a-	no a-	fo-	ra, on-	das	ver-	de-	s a-	fo-	ra,
1.162.1.1	Bem	ve-	lhi-	nho,	que es-	tá-	se ain-	da	tra-	ba-	lhas	tan-	to,
1.168.1.2	Se,	pe-	la tu-	a	por-	ta, a	lu-	z en-	tra, a	can-	tar,		
1.168.2.2	Se, em	teu-	s al-	vos	len-	çois,	co-	mo	fei-	tos	de	luar,	
1.181.4.3	Nun-	ca	pô-	de	la-	va-	r a	dor,	que a	mar-	ti-	ri-	za!
1.188.8.2	1	2	3	4	5	6	7	8	9	10	11	12	13
	Pe-	la	pri-	mei-	ra	vez,	no	di-	a	de	meu-	s a-	nos,
1.188.16.3	Tra-	zer,	de	quan-	do em	quan-	do, a	mais	tris-	te	sau-	da-	del
2.25.5.12	—“Meu	fi-	lho, o	céu	a-	zul	é u-	ma	Ca-	sa In-	fi-	ni-	ta,
2.26.22.4	fi-	lho do	seu	a-	mor,	se-	ja o	no-	me	lou-	va-	do	
2.30.3.8	O-	ra à	chu-	va in-	cle-	men-	te, ou	às	gran-	des	ge-	a-	das
4.1.73.2	1	2	3	4	5	6	7	8	9	10	11	12	13
	Na es-	tra-	da	ne-	gra e a-	troz	dos	meus	so-	nho-	s e	do-	res.
4.15.4.3	Os	de-	sí-g-	nios	fa-	tais	dos	di-	a-	s a-	zi-	a-	gos.
5.65.7.2	Bran-	ca,	de	poei-	ra	de ou-	ro, e	de a-	sa i-	rre-	qui-	e-	ta,

**Corpus SIL: tetrassílabos**

	1	2	3	4	5
1.135.5.4	—	pou-	sam	nas	flo- res.
2.35.2.3	A	do-	ce	h-	o- ra
5.125.4.1	Que	à	do-	ce	bri- sa
5.201.7.7	po-	r	e-	sses	pra- dos
5.209.5.4	Bri-	lhar	de	gló-	ria!

	1	2	3	4	5
5.222.2.1	Que	lin-	do	mi-	mo,
5.222.3.1	Zum-	be,	dan-	çan-	do

**Corpus SIL: pentassílabos**

	1	2	3	4	5	6
1.17.3.8	nos	véus	qu'es-	cu-	re-	cem.

**Corpus SIL: hexassílabos**

	1	2	3	4	5	6
5.126.56.2	a	no-	tí-	cia	cru-	el
5.126.82.4	o	seu	ben-	di-	to	a- mor!

**Corpus SIL: heptassílabos**

	1	2	3	4	5	6	7	8
1.4.2.6	no-	s a-	rre-	bóis	do O-	ri-	en-	te.
1.44.4.1	brin-	cam,	co-	lhen-	do	nos	pra-	dos
1.44.4.4	e	a-	çu-	ce-	nas	mi-	mo-	sas.
1.60.3.1	Vê	co-	mo	bri-	lham	na	rel-	va
1.71.10.3	mi-	nha	mãe...	(sau-	da-	de in-	fin-	dal)

	1	2	3	4	5	6	7	8
1.75.5.2	os	pra-	dos	re-	ver-	de-	ci-	am;
1.96.1.1	Eu	ti-	ve um	gran-	de	te-	sou-	ro,
1.96.2.11	a	ro-	sa	dos	meu-	s a-	mo-	res
2.7.8.1	Li-	ber-	da-	de,	Li-	ber-	da-	de,
2.28.2.3	Que	à-	s au-	las	pre-	si-	di-	a

	1	2	3	4	5	6	7	8
2.28.3.3	É	co-	mo um	pai —	es-	tou	ven-	do,
2.30.4.4	Se	trans-	for-	ma-	ram	por	fim!	
5.6.4.2	flo-	ri-	nhas	ru-	bra-	s e a-	zuis,	
5.14.1.4	a	des-	va-	li-	da or-	fan-	da-	de.
5.22.1.3	a	bri-	sa a-	lém	re-	pe-	ti-	a:

**Corpus SIL: heptassílabos (cont.)**

	1	2	3	4	5	6	7	8
5.26.1.5	Quem	sa-	be	das	pró-	prias	flo-	res,
5.36.1.8	e as	bor-	bo-	le-	tas	pou-	san-	do,
5.62.1.5	e a	al-	ma	ao-	s o-	lho-	s a-	sso-
5.69.1.8	pro-	cla-	ma	que	tal	sa-	ber	
5.78.1.2	as	mei-	gas	ro-	las	ge-	mi-	am,
	1	2	3	4	5	6	7	8
5.85.1.1	Há	no-	mes	cha-	os	de	gló-	ria
5.85.1.5	O	mun-	do,	mai-	ra-	vi-	lha-	do,
5.89.1.4	o	bra-	do	da	in-	de-	pen-	dên-
5.114.2.7	co-	mo	so-	nha-	res	se-	re-	nos,
5.114.4.4	e o	mun-	do	ao	tris-	te	mal-	diz!
	1	2	3	4	5	6	7	8
5.126.21.3	nem	noi-	tes	de	tem-	pes-	ta-	des,
5.126.73.2	pe-	los	lu-	ga-	res	que	vi-	a,
5.139.6.2	tão	chei-	a	d'ho-	rror,	d'es-	pi-	nhos,
5.195.4.3	Ti-	ran-	do-	se a	i-	men-	si-	da-
5.195.11.1	A-	li,	no a-	ber-	to al-	ça-	pão,	
	1	2	3	4	5	6	7	8
5.200.4.2	Tu	po-	ssuis	te-	sou-	ros	mil;	
5.200.6.2	Es-	tre-	las	que	bor-	da o	céu,	
5.226.7.2	Que	vos	per-	ten-	cem,	a	fi-	nal;
5.237.8.3	Que o	mar	so-	nhan-	do	tra-	zi-	a
5.245.7.2	Já	não	po-	dia	eu	so-	frer	
	1	2	3	4	5	6	7	8
5.245.17.2	Do en-	jô-	io a	cru-	el-	da-	ide,	

**Corpus SIL: eneassílabos**

1 2 3 4 5 6 7 8 9 10  
5.157.1.4 o-rai,;se-re-nas,;de-fé-ba-in-ha-;das.

**Corpus SIL: decassílabos**

	1	2	3	4	5	6	7	8	9	10	11
1.3.1.2	Ve-	nho hu-	mil-	de o-	fer-	tar-	te,	re-	ve-	ren-	te,
1.3.9.1	múl-	vei	à	te-	rra	vo-	sso o-	lhar	pie-	di-	so
1.8.3.3	vol-	tua, ex-	tre-	mo-	sa, ao	meu	so-	nhar	dê	vo-	da, —
1.37.3.2	a	do-	ce	quei-	xa ao	ru-	mo-	re-	jo en-	vol-	ta
1.41.2.3	e, em-	bal-	de, em-	bal-	de! —	só	cru-	éis	lem-	bran-	ças...

**Corpus SIL: decassílabos (cont.)**

1.41.6.1	1	2	3	4	5	6	7	8	9	10	11
Sor-	ve a	vio-	le-	ta a	go-	ta	do	se-	re-	no,	
1.59.6.4	1	2	3	4	5	6	7	8	9	10	11
o	vêu	da	mor-	te	so-	bre as	fa-	ces	sen-	te!	
1.62.5.4	1	2	3	4	5	6	7	8	9	10	11
ce-	les-	te	bên-	ção	que o	pra-	zer	de-	rra-	ma!	
1.64.3.1	1	2	3	4	5	6	7	8	9	10	11
Do	Céu	se-	re-	no	pe-	lo a-	zul	in-	fin-	do	
1.75.3.1	1	2	3	4	5	6	7	8	9	10	11
Ar-	can-	jo	da	Poe-	sia! —	Ô	Ser	di-	vi-	no,	
1.81.6.3	1	2	3	4	5	6	7	8	9	10	11
E e-	ssa	voz	que	no	lar	tu-	do a-	le-	gra-	va	
1.83.4.4	1	2	3	4	5	6	7	8	9	10	11
que	d'ou-	tro um	co-	ra-	ção	ter-	no	se-	pa-	ra!	
1.86.5.2	1	2	3	4	5	6	7	8	9	10	11
em	so-	nhos	d'oi-	ro, ao	des-	bro-	char	das	flo-	res,	
1.93.4.2	1	2	3	4	5	6	7	8	9	10	11
co-	mo a	que-	brar	de um	so-	nho	mau	o e-	fei-	to,	
1.104.5.4	1	2	3	4	5	6	7	8	9	10	11
a	pro-	fun-	de-	za	tua,	a-	rras-	ta,	sol-	ta...	
1.119.3.2	1	2	3	4	5	6	7	8	9	10	11
o	co-	ra-	ção	de	má-	goas	nos	tras-	pa-	ssa,	
1.127.1.3	1	2	3	4	5	6	7	8	9	10	11
A	nu-	vem	de	me-	do-	nho	tem-	po-	ral		
1.134.2.2	1	2	3	4	5	6	7	8	9	10	11
que	tem	de	mãe	a au-	rê-	o-	la	di-	vi-	na,	
1.134.4.1	1	2	3	4	5	6	7	8	9	10	11
Quem	é?	tão	lin-	da e	tão	a-	ben-	ço-	a-	da,	
1.141.3.1	1	2	3	4	5	6	7	8	9	10	11
E	pa-	ssa-	vam	se-	re-	nas	pri-	ma-	ve-	ras:	
2.9.1.21	1	2	3	4	5	6	7	8	9	10	11
Va-	le a	pe-	na	so-	fre-	r, e...	tra-	ba-	lhar!		
2.55.1.2	1	2	3	4	5	6	7	8	9	10	11
So-	bre a-	s â-	guas,	na	lu-	ta em-	bra-	ve-	ci-	da,	
3.5.1.1	1	2	3	4	5	6	7	8	9	10	11
A-	go-	ra,	de	Pi-	la-	to-	s à	pre-	sen-	ça	
3.5.4.3	1	2	3	4	5	6	7	8	9	10	11
Do	po-	vo en-	tre-	gue ao;	in-	fer-	nal	trans-	por-	te!	
3.8.4.3	1	2	3	4	5	6	7	8	9	10	11
Dei-	xa,	num	bei-	jo	de	pie-	do-	so a-	mor!		
3.9.2.2	1	2	3	4	5	6	7	8	9	10	11
En-	tre a-	me-	a-	ças	che-	ga-	lhe,	do-	ri-	do;	
3.15.2.4	1	2	3	4	5	6	7	8	9	10	11
Te-	ve,	na-	que-	la a-	tro-	z an-	si-	e-	da-	de!	
3.16.3.1	1	2	3	4	5	6	7	8	9	10	11
Vem,	Ma-	da-	le-	na,	vem;	o a-	fe-	to	san-	to	
4.9.2.1	1	2	3	4	5	6	7	8	9	10	11
La-	bo-	ra-	do-	ras	mão-	s, al-	fim,	can-	sei;		
4.12.2.2	1	2	3	4	5	6	7	8	9	10	11
E	bor-	bo-	le-	tas,	can-	tos	de a-	ves	man-	sas,	
4.13.2.3	1	2	3	4	5	6	7	8	9	10	11
Num	bri-	lho	d'ou-	ro a	ma-	ta	res-	plan-	de-	ce,	
4.16.2.2	1	2	3	4	5	6	7	8	9	10	11
fo-	ra e-	ssa	voz	do	meu	a-	mor	per-	fei-	to,	
4.17.2.2	1	2	3	4	5	6	7	8	9	10	11
sem	co-	gi-	tar	de	sú-	bi-	tas	mu-	dan-	ças;	
4.36.2.4	1	2	3	4	5	6	7	8	9	10	11
da	paz	nos	bra-	ços,	gra-	ta a-	dor-	me-	ci-	a.	
5.102.1.5	1	2	3	4	5	6	7	8	9	10	11
das	bo-	ni-	nas,	dos	lí-	rios	per-	fu-	ma-	dos,	
5.106.1.4	1	2	3	4	5	6	7	8	9	10	11
co-	mo um	a-	mi-	go	pi-	e-	do-	so es-	con-	das.	
5.117.5.4	1	2	3	4	5	6	7	8	9	10	11
te-	rá	a	tris-	te e	do-	ce	paz	do	Céu?	...	
5.117.9.3	1	2	3	4	5	6	7	8	9	10	11
— Há	no	teu	pei-	to um	co-	ra-	ção	sa-	crá-	rio —	
5.119.4.6	1	2	3	4	5	6	7	8	9	10	11
o	fi-	lhi-	nho	que	dor-	me-	des-	nu-	da-	do.	
5.123.1.4	1	2	3	4	5	6	7	8	9	10	11
as	pe-	t'las	de-	li-	ca-	das	lhe	ba-	nha-	ra.	

**Corpus SIL: decassílabos (cont.)**

	1	2	3	4	5	6	7	8	9	10	11
5.132.2.1	Plá-	ci-	da	pri-	ma-	ve-	ra!	Oh,	quan-	tas	flo- res
5.159.4.2	Ri-	so-	nho o-	á-	sis	de	mi-	mo-	sas	flo-	res,
5.170.1.1	A	ti,	que	tão	mo-	des-	to e	pe-	que-	ni-	no
5.175.4.3	Bri-	lhar	na	gran-	de U-	ni-	ver-	sal H-	is-	tó-	ria!
5.205.4.3	A	Ci-	ên-	cia, o	Po-	der	do E-	ter-	no	Deus!	

	1	2	3	4	5	6	7	8	9	10	11
5.209.2.7	De u-	ma,	mon-	ta-	nha a,	le-	ve,	som-	bra, in-	cer-	ta;

**Corpus SIL: hendecassílabos**

	1	2	3	4	5	6	7	8	9	10	11	12
1.42.4.3	fa-	na-	da-	s as	ro-	sas	de	ter-	nos	ca-	ri-	nhos,
2.7.1.3	A	te-	rra	dos	li-	vres,	de	flo-	res	ves-	ti-	da,
5.134.4.4	que	du-	ro-	s es-	pi-	nhos	res-	ta-	ram-	me en-	tão!	
5.206.5.2	A	ve-	lhi-	nha	sen-	te	bem	no	fun-	do	d'al-	ma,

**Corpus SIL: dodecassílabos**

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.40.2.3	— no	teu	rai-	o	gen-	til	me	vol-	ve a	do-	ce	cal-	ma,
1.40.3.3	oh!	lu-	a	me-	ren-	có-	ria, en-	vol-	ve	na	do-	çu-	ra
1.45.3.1	Nem	um	as-	tro	no	Céu!...	Na	te-	rra e	so-	bre a-	s á-	guas,
1.45.9.3	e	so-	bre o	ou-	ro	vil	da	ín-	fi-	ma	co-	bi-	ça,
1.112.1.26	— al-	ma,	sin-	ge-	la e	pu-	ra, e	tris-	te,	qual	vio-	le-	ta!

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.113.3.2	o	pu-	ro	co-	ra-	ção	dos	bem-	a-	ven-	tu-	ra-	dos,
1.122.18.3	sen-	tin-	do	não	po-	der	guar-	dar	lhe,	por	me-	mó-	ria,
5.210.1.4	A	te-	rra	cul-	ti-	vai,	te-	rei-	s a	re-	com-	pen-	sa!
5.220.7.3	Na	te-	rra o	lo-	do	vil —	o	lo-	do	de um	pa-	ul	
5.255.4.4	— E!	le! —	o	Deu-	s i-	mor-	tal!	E-	le! —	o	Deus	Re-	den- tor!





